



Androgen receptor aggregates studies in vitro and in a transgenic mouse model of Spinal Bulbal Muscular Atrophy (SMBA)

Bahareh Eftekharzadeh

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tdx.cat) i a través del Dipòsit Digital de la UB (diposit.ub.edu) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tdx.cat) y a través del Repositorio Digital de la UB (diposit.ub.edu) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tdx.cat) service and by the UB Digital Repository (diposit.ub.edu) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

Spinal bulbar muscular atrophy (SBMA) is a rare hereditary neuromuscular disease caused by the elongation of a (polyQ) tract in the N-terminal region of the transactivation domain of androgen receptor. Although the molecular basis of SBMA is not yet fully understood, the observation of nuclear inclusions containing AR fragments in specific tissues of SBMA patients has led to the suggestion that it is linked to AR aggregation. To characterize the molecular mechanism of this process we have investigated the structural properties of the polyQ tract present in AR as well as the early stages of its oligomerization *in vitro* by nuclear magnetic resonance spectroscopy.

To study the structural properties of AR aggregates in tissue, the Seprion ligand was used for quantification of AR aggregate load in a SBMA mouse model. A combination of atomic force microscopy, transmission electron microscopy and high-resolution microscopy was used to investigate the Seprion ligand captured aggregates from muscle and spinal cord tissue. The results indicated that aggregated structures from spinal cord extract differ remarkably from the fibrillar species isolated from muscle tissue. We found that the AR fibrils in the muscles accumulate and grow in their length as the animals age.

We hypothesize that there are differences between androgen receptor variants in muscle and spinal cord, with more N-terminally truncated AR found in muscle compared to spinal cord. We believe that truncated AR goes into aggregates and leads to AR fibrillar species in muscles. We postulate further that mutant AR97Q plays a role in the recruitment of partner interacting proteins in an age-related fashion. Our studies into the elucidation of the early stages of oligomerization indicated that the polyQ tract is partially α -helical, a propensity that increases with its length. In addition, a specific region of the N-terminus of the NTD, distinct from the polyQ tract, appears to be responsible for the intermolecular interactions that nucleate AR aggregation.

Studying the interactions between AR and the molecular chaperones Hsp40 and Hsp72 respectively, by NMR spectroscopy we found that the Hsp72 and Hsp40 both bind to ²³FQNLF²⁷ motif, whereas the Hsp40 binds also to ⁵⁴LLLLQQQ⁶¹. These findings provide a simple mechanism for the disassembly of the complex between AR and molecular chaperones required for androgen receptor function. Our findings emphasize the therapeutic potential of allosteric regulators of Hsp72 and Hsp40 and provide new insights into the role of the chaperone machinery in protein quality control in neurodegenerative diseases.

bebiochem@gmail.com

2015

Bahareh Eftekharzadeh

Ph.D Thesis

ANDROGEN RECEPTOR AGGREGATES STUDIES IN VITRO AND IN A TRANSGENIC MOUSE MODEL OF SPINAL BULBAR MUSCULAR ATROPHY (SBMA)

Bahareh Eftekharzadeh



Thesis Supervisor:
Xavier Salvatella Giralt

Laboratory of Molecular Biophysics (LMB)
Institute for Research in Biomedicine (IRB)
University of Barcelona



Universitat
de Barcelona



**Universitat de Barcelona
Facultat de Farmàcia
Institut de Recerca Biomèdica, Barcelona**

ANDROGEN RECEPTOR AGGREGATES STUDIES IN VITRO AND IN A TRANSGENIC MOUSE MODEL OF SPINAL BULBAR MUSCULAR ATROPHY (SBMA)

B a h a r e h E f t e k h a r z a d e h

2015

**UNIVERSITAT DE BARCELONA
FACULTAT DE FARMACIA
INSTITUT DE RECERCA BIOMEDICA, BARCELONA**

PROGRAMA DE DOCTORAT EN BIOMÈDICA

**ANDROGEN RECEPTOR AGGREGATES STUDIES
IN VITRO AND IN A TRANSGENIC MOUSE MODEL
OF SPINAL BULBAR MUSCULAR ATROPHY (SBMA)**

Memòria presentada per Bahareh Eftekharzadeh
per optar al títol de doctor per la Universitat de Barcelona

TESI DUTA A TERME A L'INSTITUT DE RECERCA BIOMÈDICA, BARCELONA

Director i tutor

Doctoranda

Xavier Salvatella Giralt

Bahareh Eftekharzadeh

B a h a r e h E f t e k h a r z a d e h

2015

Table of Contents

<i>Abstract of the thesis</i>	1
Chapter 1. General Introduction	3
1.1. Spinal Bulbar Muscular Atrophy (SBMA)	3
1.1.1 SBMA or Kennedy disease: a neuromuscular disorder	3
1.1.2 SBMA and involvement of AR.....	5
1.1.3 SBMA and general features of other polyQ diseases	6
1.1.4 Protein Inclusions: pathogenic or not.....	8
1.1.5 Models for toxicity of polyQ proteins in skeletal muscle cells and neurons..	9
1.1.5.1 Protein quality control (PQC) system impairment	9
1.1.5.2 Transcriptional impairment/dysregulation	9
1.1.5.3 Axonal transport impairment.....	9
1.1.5.4 Tissue specificity	10
1.1.6 SBMA treatment strategies.....	11
1.1.7 SBMA transgenic mouse models	14
1.2. Androgen receptor (AR)	17
1.2.1 AR function	17
1.2.2 AR structure	18
1.2.3 AR, a protein with a large intrinsically disordered domain	21
1.2.3.1 Intrinsically disordered proteins.....	21
1.2.4 Structural properties of AR-NTD	22
1.2.5 Importance of NMR to study IDPs.....	23
1.2.6 Use of NMR in the characterization of IDPs	24
1.3. PolyQ proteins and coiled-coil interactions	26
1.4. Introduction to heat shock proteins and to their involvement in protein misfolding diseases	29
1.4.1 Protein quality control	29
1.4.1.1 Molecular chaperones.....	29
1.4.1.2 Hsp's and their function.....	30
1.4.1.3 The Hsp70 machine	30
1.4.1.4 DNAJ proteins (Hsp40s): the co-regulators of the Hsp70 machinery system.....	31
1.4.1.5 Nucleotide exchange factors (NEF's): the co-regulators of the Hsp70 machine.....	32
1.4.2 Protein folding diseases.....	33
1.4.3 Heat shock proteins and polyQ degradation.....	33
Chapter 2. Thesis aim and objectives	37
2.1. Thesis objectives	37
Chapter 3. Materials and Methods	41
3.1. Molecular Biology methods	41
3.1.1 Recombinant Gateway technology.....	41
3.1.2 AR construct preparation.....	43
3.2. Protein expression and purification	45

3.2.1 Expression of isotopically labeled protein.....	47
3.2.2 Protein purification.....	49
3.2.2.1 Ni ²⁺ affinity chromatography.....	49
3.2.2.2 Size exclusion chromatography.....	51
3.2.2.3 TEV cleavage and Reverse Nickel.....	52
3.3. Isothermal titration calorimetry (ITC).....	52
3.3.1 Principles and instrumentation.....	52
3.4. Fluorescence polarization substrate competition (FPSC).....	53
3.5. NMR spectroscopy of Biomolecules.....	54
3.5.1 Theoretical and practical aspects.....	54
3.5.2 Basic NMR characterization of a protein: 1D and HSQC.....	54
3.5.2.1 [¹ H, ¹⁵ N]-HSQC.....	55
3.5.3 Sequential assignment of backbone resonances.....	56
3.5.3.1 ¹³ C-direct NMR experiments.....	56
3.5.3.2 4Q construct samples preparation for assignment.....	57
3.5.3.3 Resonance assignment.....	57
3.5.3.4 Temperature dependent experiments.....	58
3.5.3.5 Backbone dynamics experiments.....	58
3.5.3.6 25Q construct samples preparation for assignment.....	58
3.5.3.7 Resonance assignment.....	58
3.5.3.8 Sample preparation for interaction studies with Hsp40 and Hsp72.....	59
3.5.3.9 NMR acquisition for interaction studies between AR and Hsps.....	59
3.5.4 Secondary structure propensity.....	60
3.5.5 Relaxation and dynamics.....	61
3.5.6 ¹⁵ N Relaxation and heteronuclear NOEs.....	62
3.5.7 Proton-proton NOEs.....	62
3.5.8 Chemical shift perturbation experiments.....	62
3.6. Transgenic mouse colony maintenance.....	63
3.6.1 Re-characterization of SBMA transgenic mice.....	64
3.6.2. Transgenic mouse embryo transfer from Japan.....	64
3.6.3 Transgenic mouse breeding plan.....	66
3.7. Transgenic mouse tissue homogenate characterization.....	67
3.7.1 Biochemical characterization.....	67
3.7.1.1 Tissue dissection and homogenate preparation.....	67
3.7.1.2 Protein expression analysis in tissue homogenate (Western blot).....	67
3.7.2 RNA quantification analysis.....	68
3.7.3 Antibodies.....	68
3.8. Quantification of aggregate amount in tissue homogenate.....	69
3.8.1 Seprion ligand for the study of Protein Aggregation Diseases.....	69
3.8.2 Seprion ligand ELISA.....	69
3.9. Isolation of aggregated species from ex vivo material.....	70
3.9. 1 Seprion ligand bead capture (PAD-beads).....	70
3. 10. Cell culture and immunofluorescence staining.....	70
3.10.1 Cellular uptake and cell toxicity studies.....	71
3.11. High-resolution microscopy of AR aggregated species.....	73
3.11.1 dSTORM (Direct stochastic optical reconstruction microscopy).....	73

3.11.2 STED (Stimulated emission depletion microscopy)	75
3.12. Atomic force microscopy (AFM)	77
3.13. Transmission electron microscopy (TEM).....	78
3.14. Immunohistochemical analysis	79
3.15. Synchrotron micro-computed tomography.....	81
3.16. Statistical analysis.....	82
<i>Chapter 4. Sequence determinants of AR oligomerization in SBMA</i>.....	84
4.1. Introduction.....	84
4.2. Results	87
4.2.1 Preparation of solutions of monomeric AR fragments.....	87
4.2.2 Analysis of AR oligomerization.....	88
4.2.3 Preparation and structural characterization of a kinetically stable sample by NMR	90
4.2.4 Identification of the residues undergoing structural changes during oligomerization of 25Q.....	96
4.2.5 Predictions of the structural properties of 4Q and 25Q by bioinformatics tools.....	102
4.2.6 4Q construct temperature dependence behavior observed by NMR.....	102
4.3. Summary.....	104
<i>Chapter 5. Interaction of Hsp40 and Hsp72 with AR.....</i>	106
5.1. Introduction.....	106
5.2. Experimental procedure	108
5.3. Results	108
5.3.1 Heat shock protein72 (Hsp72) recognizes the FQNLF motif and its flanking region in AR 4Q and 25Q.....	109
5.3.2 Heat shock protein 40 (Hsp40) binds less specifically to AR.....	112
5.3.3 Presence of ATP does not significantly change the binding affinity of Hsp70	117
5.3.4 Biophysical characterization of binding affinity of Hsp40 and Hsp72 to 25Q and 4Q.....	120
5.3.5 Both Hsp72 and Hsp40 interacts with AR with 1 μ M affinity.....	121
5.3.6 Fluorescence polarization substrate competition validates the interaction between Hsp72 substrate binding domain and AR.....	123
5.3.7 The ²⁴ FQNLF ²⁸ region is necessary for the interaction of Hsp72 with AR in vivo.....	125
5.3.8 Over expression of Hsp72 and Hsp40 in C4-2 cells decreases the amount of SDS resistant AR species	126
5.4. Summary.....	127
<i>Chapter 6. Structural characterization of AR aggregates in tissues</i>.....	130
6.1. Introduction.....	130

6.2. Results	131
6.2.1 Histological and biochemical characterization of 97Q and 24Q tissue homogenates.....	131
6.2.2 PolyQ AR aggregates are SDS resistant, but not Proteinase-K (PK) resistant	134
6.2.3 PolyQ AR aggregates in muscle comprise truncated N-terminal AR; spinal cord contain mainly C-terminal species.....	135
6.2.4 Seprion ligand specifically captures AR aggregates in total tissue homogenates of SBMA transgenic animals and provides a highly quantitative assay	136
6.2.5 Seprion ligand-coated magnetic beads capture AR aggregated species for biophysical characterization.....	138
6.2.6 Fibrillar aggregates species from SBMA transgenic animals muscle tissue have identical structural properties as those generated from N-terminal truncated AR construct in vitro	140
6.2.7 AR aggregates have well-defined fibrillar structures in muscle tissue of 97Q males of SBMA.....	145
6.2.8 Only muscle tissue homogenate of SBMA transgenic mice induced cell toxicity in Neuroblastoma cells.....	147
6.2.9 AR's toxicity in muscle tissue homogenate is due to the formation of intranuclear aggregates	148
6.2.10 X-ray phase contrast characterization of muscular degeneration in young and old SBMA transgenic mice.....	152
6.2.11 High-resolution single molecule microscopy, a better way to illustrate the fibril structures	155
6.2.12 Anle138b: Could a novel oligomer modulator for disease-modifying therapy of neurodegenerative diseases such as prion and Parkinson's disease also affect SBMA mice pathology?	159
6.2.13 AR and its interacting partners.....	162
6.2.14 PolyQ expanded protein and its interacting partners.....	168
6.2.15 Identification of AR aggregates associated proteins.....	169
6.3. Summary	172
7. Discussion	174
7.1. Structural characterization of N-terminal domain of AR by Nuclear Magnetic Resonance (NMR)	174
7.2. Mapping of Hsps binding to AR	177
7.3. Characterization of AR aggregates in SBMA transgenic mouse model	179
8. Concluding remarks	182
8.1. What are the next steps toward a treatment for SBMA?	183
9. Appendix	187
9.1. AR sequence and residue numbering used in this thesis, according to Uniprot	187
9.2. HSQC assignment of AR-4Q	189

9.3. CON assignment of AR-4Q.....	193
9.4. HSQC assignment of AR-25Q.....	198
9.5. pDEST-HisMBP vector	203
9.6. Ni⁺² affinity chromatography	204
9.7. Size exclusion chromatography	205
9.8. Reverse Ni⁺² chromatography	206
9.9. Mass spectrometric analysis of 4Q and 25Q	207
10. Bibliography.....	239

زندگی کلبه دنجی ست که در نقشه خود
دو سه تا پنجره روبه خیابان دارد
گاه با خنده عجین است و گهی با گریه
گاه خشک است و گهی شرشر باران دارد
زندگی مرد بزرگیست که در بستر مرگ
به شفابخشی یک معجزه ایمان دارد
زندگی حالت بارانی چشمان تو است
که در آن قوس و قزح های فراوان دارد
زندگی آن گل سرخی ست که تو می بویی
یک سرآغاز قشنگی ست که پایان دارد
زندگی کن

جان من، سخت نگیر
رونق عمر جهان، چندصباحی گذر است
قصه بودن ما
برگی از دفتر افسانه ایه، راز بقاست
دل اگر می شکند
گل اگر می میرد
واگر باغ بخود رنگ خزان می گیرد
همه هشدار به توست
جان من، سخت نگیر
زندگی کوچ همین چلچله هاست
به همین زیبایی

Acknowledgment

Words fail me when I yearn to depict my profoundest feelings of gratitude for the people who have provided invaluable help during my thesis project.

Immeasurable appreciation and deepest gratitude for the help and support are extended primarily to my parents, *Hossein Eftekhazadeh* and *Fariba Khalil Pour* and my supportive family (*Paul Barth, Laleh Goldeck, Jens Goldeck, Arash Eftekhazadeh* and *Munisa Yasunova*) who have contributed in making this work possible.

Special thanks to *Kave Partovi*, who was always a great friend and has been supporting me during the last 10 years in both my scientific and non-scientific life.

Moreover, the help and support are extended to the following persons who in one way or another have helped me through this journey.

My supervisor of thesis; *Prof. Xavier Salvatella*, for his support, advice, guidance, intelligence, valuable comments, suggestions and last but not least, the freedom he gave me to develop my own thoughts and follow my ideas while giving me the direction how to always stay in the highest scientific path. His help in analyzing the data and preparing manuscripts has been highly valuable and is really appreciated.

Prof. Joan Guinovart; for his time and effort in supporting my science and applications for different competitive events and including me in important knowledge transfer possibilities which were rare chances given to me in my life.

Prof. Angelo Azzi; for his unlimited support during my whole scientific career path during the past 7-8 years by showing me the opportunities and helping me with decision making in different steps of my life.

Prof. Christopher Dobson; for giving me the main idea of protein misfolding and their involvement in neurodegenerative diseases when I was twenty years old. He was the reason for me to be where I am now and he was the reason of my decision to invest my life to study neurodegenerative diseases.

Dr. Daniel Nietlispach; for his unlimited time and effort in checking my English in my abstracts and my short stay grant applications. I thank him for his scientific support and ideas during my whole PhD project. His direct involvement in the project and consequently his helps and interesting ideas helped me in better and a deeper understanding of the scientific questions, projects, results and data.

Dr. Gabriele Kaminski Schiele; for her help, time and support during my short stay in Cambridge, UK. The fruitful discussions with Gabi made my best time in Cambridge. Her words of encouragements helped me stick to my scientific career plan and reflect on how to be a strong woman in science, as she is.

Prof. Michele Vendruscolo; for the very interesting discussions we had during my short stay in Cambridge University. He also helped me a lot with very important decision making in my scientific career.

Prof. Clemens Kaminski; for giving me the opportunity to work in his laboratory in the Department of Chemical Engineering at the University of Cambridge, UK; thanks for being so positive and enthusiastic and motivating with your words.

Dr. Marga Gairi; for being always there with her special kind supportive soul, who sustained me mentally during the time when I was writing my thesis and for being one of my best few friends in the Institute for research in Biomedicine (IRB) and in Barcelona.

Prof. Isabella Felli; for being such an amazing friend, scientist and collaborator. I give my special thanks to Isabella as she, beside my supervisor, was one of the main reasons for the success of the nicest part of my whole PhD work, which I performed in her laboratory.

Prof. Jason Gestwicki; for the very nice collaboration with us. This collaboration brought much new understanding in the field of polyQ diseases and chaperons.

Prof. Gen Sobue; for his kindness to host me for two weeks in Nagoya Medical School, he provided me in Japan with an amazing work experience which was not comparable with that of any other place.

Prof. Hiroaki Adachi; for his time in teaching me all details of the work with transgenic animals of Kennedy's disease, which were generated in Prof. Gen Sobue laboratory. He was a great host during my time in Japan.

Dr. Claudio Di Sanza; for being such a great friend and collaborator during the last few months. He did lots of work for making my project successful. His great help shows clearly in my thesis.

Giulio Chiesa; for being my first and best friend in Barcelona, in the lab and out of the lab. Giulio spent most of the time with me in the same lab and helped me a lot from the first day I started my PhD. His "relatively" smart ideas brought us lots of fun and good moods and memories in late hours of weekends in the lab.

Alessandro Piai; for being one of those few trustworthy friends and collaborators who spent many late nights with me in CERM NMR facility in Florence to do more experiments and analyze more data and who was very patient in teaching me how to make amino acid assignments in a protein and gain a practical view of NMR experiments.

Joan Miquel Valverde; for his help during the last four years in a very small but important part of my lab life, he put time to translate Spanish to English for me; he made many phone calls on my behalf while doing a perfect technician job at the same time.

Dr. Jesús García; for his great help in the analysis of NMR data and sharing his ideas with me.

Ela Szulc; for being such a great friend in the last months of my stay in Barcelona, inside and outside of the lab, she always kept her cute smiley face, which always motivated me more to overcome the challenges of finishing the PhD.

Busra Topal; For being such a great friend in the last months of my stay in the lab; being from Turkey, we shared many similar interests, which made her even more special as a friend for me.

Dr. Víctor Buzón; for his time and help during these last months and also for being a friend with a special personality.

Dr. Jesús Ureña; for his presence in my thesis advisory committee and all his supports. He has been always a positive person who motivated me to continue with lots of energy after each thesis advisory committee.

Dr. Nàtalia Carulla; for her support and presence in my thesis advisory committee and her nice ideas in all of the committee meetings. The meeting on Alzheimer's disease research she organized in Barcelona made it possible for me to finally meet Prof. Dobson and talk to him for the first time.

Dr. Daniel Lavery; for his advices during the last months on my interviews and my Postdoc position plans.

Dr. Romain Laine; who patiently taught me high-resolution microscopy and greatly helped me in my high-resolution microscopy studies during my short stay at the University of Cambridge.

Dr. Tobias Jochum; for hosting me in Karlsruhe Institute of Technology (KIT) for three months in Germany. His contribution, in some parts of my *in vivo* studies, can be well recognized in my thesis.

Abstract of the thesis

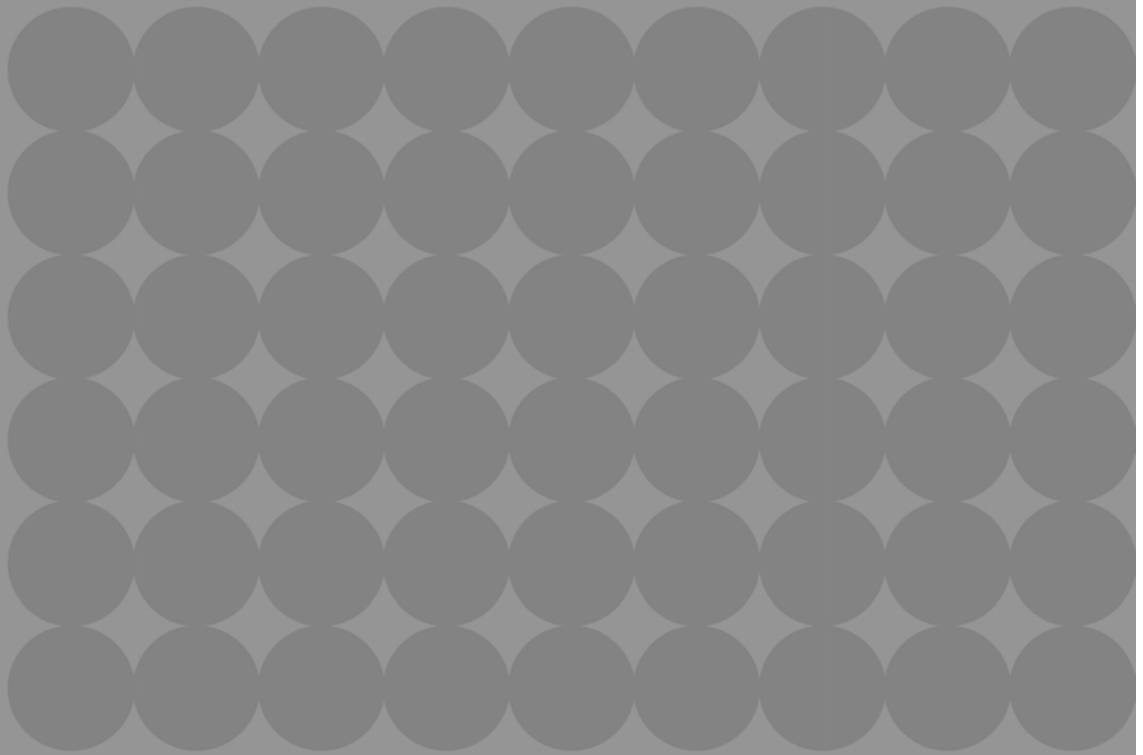
Spinal bulbar muscular atrophy (SBMA) is a rare hereditary neuromuscular disease caused by the elongation of a polymorphic polyglutamine (polyQ) tract in the N-terminal region of the transactivation domain (NTD) of androgen receptor (AR). Although the molecular basis of SBMA is not yet fully understood, the observation of nuclear inclusions containing AR fragments in specific tissues of SBMA patients has led to the suggestion that it is linked to AR aggregation. To characterize the molecular mechanism of this process we have investigated the structural properties of the polyQ tract present in AR as well as the early stages of its oligomerization *in vitro* by nuclear magnetic resonance (NMR) spectroscopy in solution.

To study the structural property of the AR aggregates in tissue, the Seprion ligand was used for quantification of AR aggregate load in a SBMA mouse model. A combination of atomic force microscopy (AFM), transmission electron microscopy (TEM) and high-resolution microscopy was used to investigate the Seprion ligand captured aggregates from muscle and spinal cord tissue. The results indicated that aggregated structures from spinal cord extract differ remarkably from the fibrillar species isolated from muscle tissue. We found that the AR fibrils in the muscles accumulate and increase their length as the animals age.

Our results indicate that there are differences between the AR variants in muscle and spinal cord, with more N-terminally truncated AR found in muscle compared to spinal cord. We propose that truncated AR forms aggregates and leads to AR fibrillar species in muscles and that mutant AR97Q plays a role in the recruitment of proteins in an age-related fashion.

Our studies into the elucidation of the early stages of oligomerization indicated that the polyQ tract is partially α -helical, a propensity that increases with its length. In addition, a specific region of the N-terminus of the NTD, distinct from the polyQ tract, appears to be responsible for the inter-molecular interactions that nucleate AR aggregation.

Studying the interactions between AR and the molecular chaperones Hsp40 and Hsp72 respectively, by solution NMR spectroscopy we found that the Hsp72 and Hsp40 both bind to ²³FQNL²⁷ motif, whereas the Hsp40 binds additionally to ⁵⁴LLLLQQQ⁶¹. These findings provide a simple mechanism for the disassembly of the complex between AR and molecular chaperones required for AR function and emphasize the therapeutic potential of allosteric regulators of Hsp72 and Hsp40 and provide new insights into the role of the chaperone machinery in protein quality control in neurodegenerative diseases.



Part I.

I n t r o d u c t i o n

Chapter 1. General Introduction

- 1.1. Spinal Bulbar Muscular Atrophy (SBMA)
- 1.2. Androgen receptor (AR)
- 1.3. PolyQ proteins and coiled-coil interaction
- 1.4. Heat shock proteins and polyQ diseases

Chapter 1. General Introduction

1.1. Spinal Bulbar Muscular Atrophy (SBMA)

1.1.1 SBMA or Kennedy disease: a neuromuscular disorder

Spinal and bulbar muscular atrophy (SBMA, Kennedy's disease) is an X-linked inherited neuromuscular disorder characterized by adult onset proximal muscle weakness due to lower motor neuron degeneration. SBMA patients also display signs of androgen insensitivity, including gynecomastia, reduced fertility and testicular atrophy (1, 2). This finding, together with the X-linked inheritance, led to analysis of the androgen receptor (AR) gene as the potential cause of SBMA. While a CAG repeat in the first exon of the AR gene varies in length from 5–34 triplets in normal individuals, SBMA patients were found to harbor repeats ranging from 37–66 CAG repeats (3, 4). For decades, researches into the basis of neurological disease have focused upon the contribution of neuronal dysfunction to disease pathogenesis. However, over the last 15 years, there has been a growing appreciation of the importance of non-neuronal cells in maintaining neuron function and contributing to neurological disease pathogenesis (5).

SBMA patients with neuromuscular symptoms, very often are misdiagnosed for amyotrophic lateral sclerosis (ALS). The clinical features of both diseases are very similar, but the progression in ALS is much faster than SBMA. Recent researches in the field of neuromuscular disease have provided convincing evidence of the involvement of motor neurons in ALS, while more muscle problems are diagnosed in SBMA (6, 7).

Skeletal muscle is a major source of trophic support for innervating motor neurons and has been shown to contribute not only to neuron survival during development, but also to synaptic activity and axonal function (8). SBMA patients often exhibit features of myopathy, as progressive muscle weakness occurs in the context of elevated serum creatine kinase levels (9). Muscle biopsies of SBMA patients reveal mixed pathological findings, with both myopathy and neurogenic atrophy features (10) and presence of intranuclear inclusions (Fig. 1-1).

As shown in figure 1-2, in 2002 Christopher A Ross published a model (11) for the mechanism of toxicity of AR with an expanded polyQ in SBMA. The mechanisms of neurodegeneration in the CAG repeat polyQ diseases, including Spinal and Bulbar Muscular Atrophy (SBMA), Huntington's disease (HD), DentatoRubral and PallidoLuisian Atrophy (DRPLA), and Spino-Cerebellar Ataxia (SCA) have been controversial and there is still no clear mechanism which can explain the whole disease pathology. Proposed mechanisms have included activation of Caspases (12, 13) or other triggers of apoptosis, mitochondrial or metabolic toxicity and interference with gene transcription.

One of the main features of SBMA pathology is the presence of nuclear inclusions of AR in spinal cord and skeletal muscle cells (Fig. 1-1) (14). It was demonstrated in 2005 (15) that diffuse nuclear accumulations of mutant AR frequently and extensively were distributed in CNS and even in visceral organs and this was not limited to spinal cord and muscles. But it was noted that spinal cord and skeletal muscles comprised mainly nuclear inclusions and not the cytoplasmic variants of AR. Recent studies have revealed that although different variants of AR aggregates, intracellular, in cytoplasm or intranuclear exist, mainly the intranuclear ones show toxic properties.

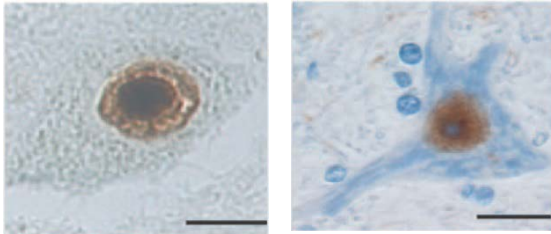
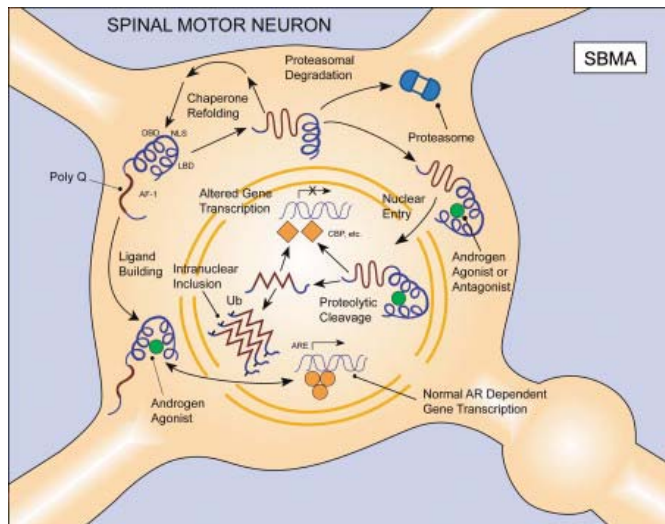


Figure 1-1: a) Nuclear inclusions of AR. AR aggregates were stained using 1C2 antibody in anterior horn neurons. b) Web-like pattern of nuclei in presence of aggregates of AR in anterior horn neurons. Adapted from (15).

Figure 1-2: Model for SBMA cellular pathogenesis. The AR is cytoplasmic (and bound to Hsps, not shown) in its basal state. It consists of several domains, including an N-terminal transcriptional regulatory region (AF-1), a zinc finger containing DNA binding region (DBD), a hinge region with the nuclear localization signal (NLS), and a C-terminal ligand binding domain (LBD). Upon ligand binding, there is a conformational change that results in homo-dimerization and dissociation from Hsps



(not shown) and exposure of the NLS, resulting in nuclear translocation. In the nucleus, agonist binding results in association with androgen response elements (ARE), which leads to the recruitment of co-activators and the activation of gene transcription. By contrast, antagonist binding results in nuclear translocation without targeting to the androgen response elements or recruitment of co-activators and thus, no gene transcription. The pathogenesis of SBMA may result from targeting of the mutant receptor to the nucleus. The expanded polyQ tract causes an altered conformation of the protein. Upon binding of either agonist or antagonist, the receptor translocates to the nucleus. After proteolytic cleavage (either in the nucleus or the cytoplasm), the polyQ stretch assumes an altered conformation, leading to aggregation and the formation of intranuclear inclusions. The mutation may confer a gain of a novel toxic property on the AR, such as abnormal interactions with CBP, leading to a loss of neuronal survival signaling. Adapted from (11).

1.1.2 SBMA and involvement of AR

AR is a member of the steroid hormone receptor family of ligand-dependent nuclear receptors. AR functions include gene expression via action as a DNA-binding transcription factor, cell cycle/proliferation regulation, cell-to-cell signaling and intracellular signal transduction, leading to the regulation of biological processes such as development, cellular proliferation, differentiation and apoptosis. The presence of AR is required in the development of the male phenotype including muscle growth and male patterning of the brain. Its gene is located on the long arm of the X chromosome, Xq11-12, and consists of 8 exons (Fig. 1-3).

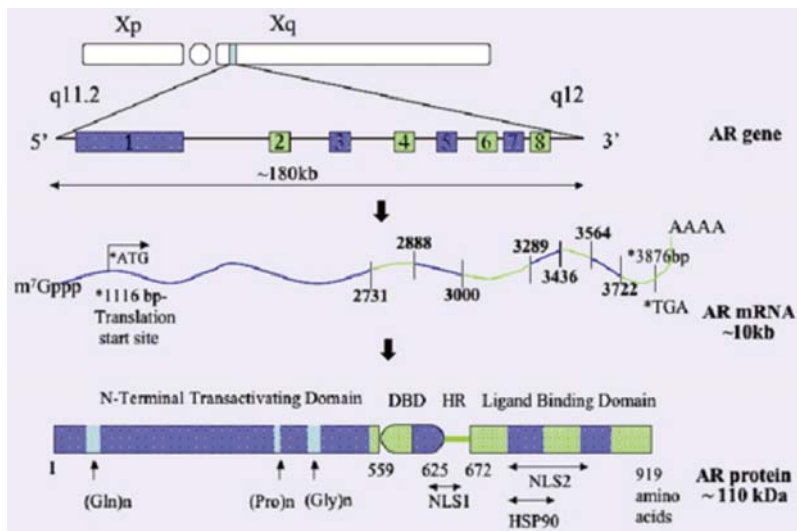


Figure 1-3: Schematic representation of the human AR gene, AR mRNA and the AR protein with indicated regulatory regions, adapted from (16).

In 1991 X-linked spinal bulbar muscular atrophy or Kennedy's disease was introduced by Kenneth Fischbeck and his colleagues (4) as an adult onset form of motor neuron disease, which may be associated with signs of androgen insensitivity. They investigated whether the AR gene on the proximal long arm of the X-chromosome was a candidate gene for this disease. They found AR gene mutations with increased size of a polymorphic tandem CAG repeat in the coding region. These amplified repeats were clearly associated with the disease, being present in 35 unrelated patients and none of the 75 controls and showed that the mutations were segregated with the disease in 15 families, and all the genetic probability calculations made the association unlikely to be due to linkage disequilibrium. Therefore, they concluded that the enlargement of the CAG repeat in the AR gene is the most likely cause of this neuromuscular disorder (4).

In SBMA, the expansion of an unstable CAG repeat in the coding region of the AR gene causes the protein to aggregate and to lose its functionality. The number of CAGs is inversely correlated with the age of onset of the disease.

Intergenerational CAG repeat expansion is observed predominantly in parental rather than maternal transmissions, suggesting that particular instability of the CAG repeat occurs in spermatogenesis (17).

1.1.3 SBMA and general features of other polyQ diseases

As CAG encodes the amino acid glutamine (Q), SBMA was the first disorder identified to result from expansion of a polyQ repeat tract. Eight other inherited neurodegenerative disorders were subsequently found to be caused by expanded CAG repeats (18). PolyQ disorders are diseases of the central nervous system with involvement of muscle disorders, caused by a CAG triplet expansion in the affected genes, that leads to protein aggregation, in which muscular and neuronal inclusion formation by the mutant protein hallmarks the disease (19–21). The 9 neurodegenerative disorders, caused by these usually unstable trinucleotide tracts are: spinal bulbar muscular atrophy (SBMA, Kennedy's disease), Huntington disease (HD), DRPLA (dentatorubropallidoluysian atrophy) and the spinocerebellar ataxia's (SCA1, 2, 3, 6, 7, and 17) (19). These neurological disorders are autosomal dominant, except for SBMA, which is an X-linked recessive disease (21–23). The disorders usually start in midlife and slowly progress, often leading to wheel chair dependency in 10-20 years and causing a progressive increasing of neuronal dysfunction and neuronal cell loss and muscle dysfunction (20). Table 1-1 summarizes the main characteristics of each disorder. Tables 1-2 and 1-3 summarize the affected brain regions in every disorder, and the corresponding symptoms.

The age of onset (AO) of all polyQ diseases is inversely correlated with the CAG repeat length: the longer the repeat, the earlier the onset and the severity of the disease. The expanded repeats are unstable and tend to expand further over generations, especially when the father is affected, leading to earlier age of onset and more severe phenotypes (24, 25). On average, the AO of all polyQ disorders is determined by the length of the CAG repeat (24, 26); this implies that the rest of the variation is likely to be determined by environmental and other genetic factors which are thought to modify the course of the disease and can form the basis for a therapy delaying the AO (27) or a phenotypes modifying therapy.

Table 1-1 summarizes the characteristics of each disorder. The mutated gene encoding the affected protein, the limit for the CAG repeat in normal (CAG<) and disease (CAG>) conditions, the average age of onset (AO): the age of the patient when the first signs of disease were observed, (CAG-AO) the percentage in which the CAG repeat length determines the age of onset, and the localization of the inclusions (NI= nuclear inclusions, CI=cytoplasmic inclusions). Adapted from (21).

Disorder	Gene	CAG <	CAG >	AO	CAG-AO	Inclusions
SBMA	Androgen receptor	9-36	38-62	40	60%	NI
HD	HD gene	1-34	36-121	40	50%	NI and CI
SCA1	Ataxin 1	6-44	39-82	35	66%	NI
SCA2	Ataxin 2	15-24	32- >100	35	73%	CI
SCA3	Ataxin 3	13-36	55-84	40	45-60%	NI (a few CI)
SCA6	CACNA1A	4-18	20-33	50	44%	CI
SCA7	Ataxin 7	4-35	37-306	30	75%	NI
SCA17	TATA binding protein	25-42	47-63	30	42%	NI
DRPLA	Atrophin-1	7-34	49-88	30	50-68%	NI

Table 1-1: Characteristics of polyQ disorders (Adapted from 21)

Disorder	Affected regions
SBMA	Anterior horn, dorsal root ganglia
HD	Striatum, cortex
SCA1	Cerebellum, dentate nucleus, brain stem
SCA2	Cerebellum, brain stem, fronto-temporal lobes
SCA3	Cerebellum, basal ganglia, brain stem, spinal cord
SCA6	Cerebellum, dentate nucleus, inferior olive
SCA7	Cerebellum, brain stem, macula, visual cortex
SCA17	Cerebellum, cortex, basal ganglia
DRPLA	Cerebellum, cortex, basal ganglia

Table 1-2: Affected brain regions in polyQ disorders (Adapted from 21)

Disorder	Main clinical features
SBMA	Motor weakness, swelling, gynecomastia, decreased fertility
HD	Chorea, dystonia, cognitive deficits, psychiatric problems
SCA1	Ataxia, dysarthria, spasticity, cognitive impairments
SCA2	Ataxia, polyneuropathy, slow saccades
SCA3	Ataxia, parkinsonism, spasticity, extra pyramidal signs, polyneuropathy
SCA6	Ataxia, dysarthria, nystagmus
SCA7	Ataxia, blindness, dysarthria, extra pyramidal signs
SCA17	Ataxia, cognitive decline, seizures, and psychiatric problems
DRPLA	Ataxia, seizures, choreoathetosis, dementia

Table 1-3: Main clinical features of polyQ disorders (Adapted from 21)

Table 1-3 summarizes the main symptoms of polyQ disorders. Gynecomastia: Abnormal development of large mammary glands in males. Chorea: Dance like abnormal involuntary movements. Choreoathetosis: Occurrence of involuntary movements. Dystonia: repetitive movements or abnormal postures by muscle contractions. Dysarthria: poor articulation. Nystagmus: Involuntary eye

movement. Polyneuropathy: simultaneous malfunction of peripheral nerves. Adapted from (21).

As stated above, polyQ disorders are a heterogeneous group of neurodegenerative diseases caused by a CAG expansion in unrelated proteins, leading to the formation of insoluble, fibrillar aggregates in muscles and neurons containing elements of protein quality controls like molecular chaperons, ubiquitin, proteasomes and transcription factors (28). These resemblance indicate that some of the mechanism of pathogenesis that connect the disorders might be conserved, but each with specific elements in each disease.

Despite extensive research, the molecular mechanism of polyQ aggregation and its cellular toxicity is not well explained yet. Moreover, the data on the role of polyQ inclusions' toxicity is contradicting, making it difficult to determine whether inclusions are protective, toxic or both (29–31). In the following, various general mechanisms in polyQ disease progression are discussed.

1.1.4 Protein Inclusions: pathogenic or not

Similar to other neurodegenerative disease, one of the major hallmarks of all polyQ diseases is the large protein inclusion found both in nucleus and the cytosol, comprising amyloid structures. Evidence for the toxicity of inclusions is based on the observation that inclusion formation in cultured cells correlates with cell death (32). Moreover, in animals the amount of polyQ aggregates correlates with the disease progression (33), increases when the disease progresses and is reversed when the expression of the mutant protein is turned off (34). Molecules able to interfere with the formation of aggregates in Alzheimer's disease or Parkinson's disease, like Congo red, Thioflavine and green tea have reduced toxicity in cellular and animal models (35–37); also Hsps inhibiting protein aggregation decrease the polyQ toxicity (38–40). However, other evidence has suggested that these aggregated entities may not be the primary toxic events causing the manifestations of the disease (41). In human HD post-mortem brains, the presence of inclusions is not directly correlated with neuronal cell loss, as the highest numbers of neurons containing inclusions are found in non-degenerating neurons (29, 42), and degenerated areas sometimes lack inclusions. Moreover the HD shortstop mouse model, which expresses a short fragment of human Huntingtin (Htt) with an expanded polyQ repeat shows widespread Htt inclusions in tissues with no clear pathology (31). Also in experimental *in vitro* cellular models, inclusions are often not correlated with cell death (41, 43).

In spite of these apparent controversies, oligomerization does seem to be crucial for the pathogenesis of polyQ diseases (37). It is assumed that early intermediates in the aggregation pathway (monomers, disordered oligomeric fragments or aggregates) are the most toxic species (44) and that the formation of inclusions represents a last attempt of the cells to protect themselves against these toxic

species. However, as I will show the aggregated species of AR are highly important in molecular mechanism of SBMA disease.

1.1.5 Models for toxicity of polyQ proteins in skeletal muscle cells and neurons

How the toxic species, whether monomeric or higher order aggregates, initiate the cascade of pathogenic protein-protein interactions that trigger neuronal dysfunction is still not well understood. A few of the hypothesis are described in the following.

1.1.5.1 Protein quality control (PQC) system impairment

Molecular chaperones and Ubiquitin-proteasome system (UPS) components have been found to be associated with polyQ inclusions in which they may be sequestered, leading to loss of function (28, 34). It is known that polyQ proteins sequester other proteins with specific sequence properties, either rich in Q/N (Gln/Asn) or small elongations of Gln tracts (45–47). These might be one of the possibilities for impairment of protein quality control. In addition, direct clogging up the proteasome may occur during the digestion of soluble expanded polyQ proteins (34, 48, 49). These events, together with a normal increase in metastable proteins during ageing (43, 50) and an age-dependent decline in these PQC systems, will impair protein homeostasis leading to a self-perpetuating and progressive global cellular dysfunction which ultimately causes cell death.

1.1.5.2 Transcriptional impairment/dysregulation

It has been shown by microarray gene expression profiling, that in presence of polyQ proteins, the expression of many genes is altered, some of which are protein context dependent, but others seem to be generally dysregulated in all polyQ diseases (51–53) As previously mentioned, transcriptional dysregulation and toxicity could arise from the recruitment of other polyQ containing proteins during the aggregation process into the aggregate that hereby may lose their physiological function (47, 54–56). Interestingly, a number of transcriptional regulators become trapped in nuclear inclusions (54, 57–60). For example, the transcriptional co-activator CREB-binding protein (CBP) with a polyQ tract is inactivated in polyQ expressing cells and toxicity and can be partially prohibited by its re-introduction (61–64).

1.1.5.3 Axonal transport impairment

Similar to most of the neurodegenerative diseases with presence of large amyloid proteins intracellular or extracellular of the cells (65, 66), perturbations in transport pathways within the long, narrow-caliber axons are likely to be an early event also in (some) polyQ disorders, causing neuronal dysfunction, neuronal death and eventually muscle cells impairment (57). These axonal transport disruptions have been observed in *Drosophila* HD models, which are characterized

by the presence of aggregate in the cytoplasm, sequestering other polyQ containing proteins in the cytoplasm leading to the disruption of the axonal transport and an accumulation of aggregates at synapses (67). Moreover, this phenomenon was also observed in HD mice where degenerated axons were observed with Htt aggregates associated with degenerated mitochondria (68), and dystrophic neurites, which are extracellular structures containing axonal processes. Consistently, in brains of e.g. HD and SCA3 patients' axonal inclusions have been also detected (69, 70).

The above-mentioned toxicity pathways are not mutually exclusive (e.g. transcriptional dysregulation could be failure of PQC systems to chaperone transcription factors) and they may occur in parallel or sequentially with other events. This may also be relevant to the discussion of nuclear versus cytoplasmic aggregation and their role in polyQ proteins toxicity (15). The finding that nearly all polyQ disorders display predominantly nuclear inclusions prompted the hypothesis that nuclear aggregation is required and sufficient for pathogenesis (15, 17). Indeed, in many experimental models, modulating the localization of polyQ proteins has shown effects on the disease manifestation. As an example, preventing nuclear entry attenuated nuclear localization and enhanced cellular toxicity (15, 41, 42) and disease onset in mice. Requirement of nuclear aggregation can also be easily explained in terms of the transcriptional dysregulation hypothesis (71) and with the findings that especially due to the absence of most of the molecular chaperons or their low expression level, the nucleus is an aggregation-prone environment (72), it might be also because systems like autophagy are not operational in nuclei (73) and therefore the nuclear fragments escape the cytoplasmic quality control (74). Indeed, in several polyQ diseases like SCA2 (75), SCA6, HD (69) and even recently in SCA3 (70) protein aggregates have been identified exclusively or additionally in the cytoplasm. These cytoplasmic inclusions may impair axonal transport and contribute to degeneration of skeletal muscles and nerve cells (57). Moreover, in some cases, the observed toxicity might be independent of the formation of aggregates but might be caused by the destabilization of the microtubules for examples by full-length mutant Htt (76).

It is important to note that early events in the progression of the disease should be also taken into account. Studying the late stage characteristics of amyloid diseases may clarify many aspects of the disease, however, a detailed look into the whole procedure of disease progression from pre-onset to postmortem is mandatory for polyQ diseases field.

1.1.5.4 Tissue specificity

Despite the widespread expression of all of the mutant polyQ proteins throughout the body, in most of these diseases, with the exception of SBMA, the central nervous system is particularly vulnerable and neurodegeneration is causative for the devastating (19, 77, 78). As neurons are post mitotic cells, their ability to cope

with misfolded proteins may be different from that of other tissue which can proliferate and regenerate. Identical reasons for muscle tissue vulnerability also exist. Although neurons are a distinct target of polyQ-mediated toxicity, pathological changes in non-neuronal cells also play a role in the development of neurodegeneration (79, 80). Special features of skeletal muscle tissues make them highly susceptible to the presence of aggregates, not only due to their chemical and structural properties, but also their physical features like size and stability. Muscle tissue highly compact with a unique cytoplasmic environment that may not withstand the formation of large aggregates of AR. Large aggregates could work as physical hindrance and therefore result in loss of elasticity in muscle cells.

Myopathic changes in muscle biopsy specimens and elevated serum creatine kinase levels in patients with SBMA imply a direct involvement of the skeletal muscle (81). Morphologic and molecular changes associated with myopathy are detectable before the initiation of neurodegeneration in a knock-in mouse model of SBMA (82, 83). Additionally, the targeted overexpression of Wt rat AR in skeletal muscle induces motor axon loss in mice. These observations suggest that AR-mediated myopathy contributes to non-cell autonomous degeneration of spinal motor neurons (84).

Asymmetric segregation of accumulated protein damage during cell division (85) seems to serve the stem cell rejuvenation and preserve highly regenerative tissues. This may explain the hypersensitivity of the neurons and muscle tissue to protein misfolding. But also within post-mitotic regions of central nervous system, large differences in sensitivity toward degeneration seem to present with each polyQ disorder showing degeneration in different areas of the central nervous system (20).

This indicates that function and/or metabolism of the affected protein and the protein context outside the CAG-repeat also plays an important role in cell-type and area-specific degeneration. Therefore, despite the various similarities between the diverse polyQ diseases, there are substantial involved regional differences between in different polyQ disorders.

1.1.6 SBMA treatment strategies

During the two decades since the discovery of the AR gene mutation in SBMA, basic and clinical research lead to a deep understanding of the disease manifestations and pathophysiology (Fig. 1-4). However, despite positive results in animal studies, no therapy has proven to be effective in clinical trials, suggesting a need to elucidate the entire disease mechanism, the early initiation of therapeutic intervention and sensitive outcome measures need to evaluate drug effects. In addition, the efficacy of non-pharmacological approaches, including physical therapy, should also be rigorously tested. Further basic and clinical studies are needed to elucidate the pathogenesis of SBMA and develop effective therapies.

The ubiquitin-proteasome system and autophagy constitute the most important cellular defense machineries against the accumulation of misfolded proteins. Treatment with 17-allylamino-17 demethoxygeldanamycin, a potent Hsp90 inhibitor, facilitates the proteasomal degradation of pathogenic AR and thereby ameliorates the motor deficits in a mouse model of SBMA (86). Pathogenic AR retained in the cytoplasm is subjected to protein degradation via the autophagic/lysosomal pathway. Pharmacological activators of autophagy, including Rapamycin, also suppress neuronal damage in cellular and *Drosophila* models of SBMA (87). Hsps mitigate polyQ-mediated cytotoxicity by refolding and solubilizing the pathogenic proteins. Overexpression of the inducible form of human Hsp70 facilitates the proteasomal degradation of abnormal AR protein and markedly ameliorates the symptomatic and histopathological phenotypes in SBMA mice (39). Favorable effects are also exerted by a pharmacological induction of molecular chaperones.

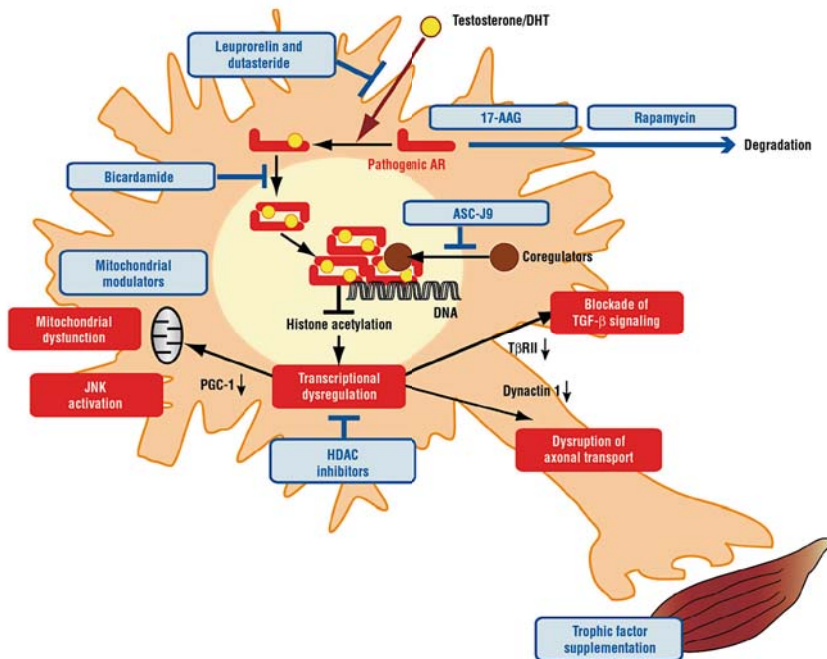


Figure 1-4: Pathophysiology and potential therapies for SBMA. Ligand-dependent nuclear accumulation of the pathogenic AR has been construed as the initial step in the neurodegeneration process of SBMA, however, DNA binding, interdomain N/C interaction and recruitment of co-regulators are also required for AR toxicity. PolyQ-expanded AR dysregulates the transcription of several genes including peroxisome proliferator-activated receptor γ coactivator 1 (PGC-1), dynactin 1, and type II transforming growth factor- β (TGF- β) receptor (T β RII), leading to mitochondrial dysfunction, axonal transport defects, and disruption of TGF- β signaling. Pathogenic AR also activates the c-Jun N-terminal kinase (JNK) pathway, culminating in apoptosis. Several potential therapies for SBMA have emerged from animal studies, e.g. hormonal interventions (leuprorelin, dutasteride, and bicardamide), enhancers of protein degradation (rapamycin and 17-allylamino-17-demethoxygeldanamycin [17-AAG]), an inhibitor of co-regulator recruitment (5-hydroxy-1,7-bis[3,4-dimethoxyphenyl]-1,4,6-heptatrien-3-one [ASC-J9]), mitochondrial modulators, histone deacetylase (HDAC) inhibitors, and trophic factors (vascular endothelial growth factor and insulin-like growth factor 1). DHT indicates dihydrotestosterone. Adapted from (9).

PolyQ AR is expressed by both lower motor neurons and skeletal muscles. Although viewed as a motor neuronopathy, data from patients and mouse models suggests that muscle contributes to disease pathogenesis. In recent studies, this hypothesis has been tested using AR113Q knock-in and human bacterial artificial chromosome/clone (BAC) transgenic mice that express the full-length polyQ-AR and display androgen-dependent weakness, muscle atrophy, and early death (88). Antisense oligonucleotides were developed that suppressed AR gene expression in the periphery but not the CNS after subcutaneous administration (Fig. 1-5)(89). Suppression of polyQ-AR in the periphery, rescued deficits in muscle weight, fiber size, and grip strength, reversed changes in muscle gene expression, and extended the lifespan of mutant males. The results concluded that polyQ-AR expression in the periphery is an important contributor to pathology in SBMA mice and that peripheral administration of therapeutics should be explored for SBMA patients.

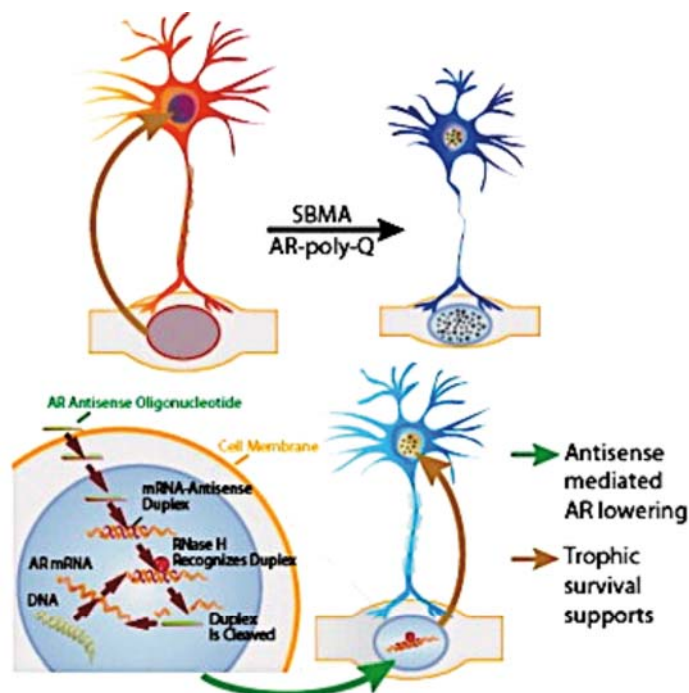


Figure 1-5: Peripheral AR gene suppression rescues disease in mouse models of SBMA. AR-targeted antisense oligonucleotides suppressed gene expression in mice. Subcutaneous delivery suppressed AR gene expression in the periphery but not the CNS. Subcutaneous administration rescued disease in two mouse models of SBMA. Peripherally expressed polyQ AR contributes to disease and is a therapeutic target. Adapted from (90).

Miyazaki *et al.* in a Nature Medicine paper (91) showed that SBMA can be treated by modulating the organism's natural biology via a specific miRNA (fig. 1-6). Initially, they observed that five miRNAs (miR-196a, miR-196b, miR-496, miR-323-3p and miR-29b) were upregulated in the spinal cords of SBMA transgenic mice

and individuals with SBMA. By overexpressing one of these miRNAs both the mutant CAG-expanded AR transcript and the mutant protein in SBMA mice were down regulated, leading to an improved neurology, body weight and survival. This notable broad-spectrum improvement in the disease phenotype is probably due to the mode of delivery of the therapeutic miR-196a using a recombinant adeno-associated virus (AAV) vector. The AAV vector was efficiently transported via the blood from skeletal muscle to the brain and the motor neurons of the spinal cord.

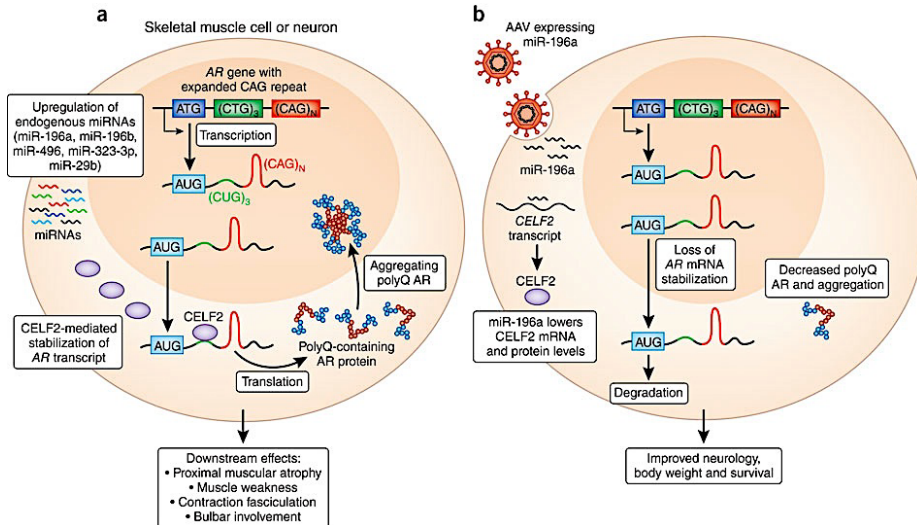


Figure 1-6: miR-196a, along with four other miRNAs, is upregulated in SBMA. The authors found that miR-196a destabilizes the CELF2 transcript, which encodes an RNA-binding protein, CELF2 that normally stabilizes the AR transcript, presumably by binding to a short CUG tract upstream of the CAG repeat in the AR transcript. Thus, administration of AAV overexpressing miR-196a targets CELF2, leading to the degradation of the AR transcript and reduced translation of the AR protein, and improves the phenotype of transgenic mice modeling SBMA. Adapted from (91).

1.1.7 SBMA transgenic mouse models

SBMA is a gender-specific disease only males are fully affected. Females, even if homozygous for the mutation, have few if any symptoms (23). In both transgenic and knock-in mouse models of SBMA, male but not female mice expressing mutant AR develop full disease manifestations (92, 93). Importantly, reduction of testosterone levels in male mice ameliorates disease manifestations, suggesting a potential therapy for SBMA (94). Indeed, a phase 2 clinical trials shows the benefits of androgen deprivation by leuprorelin acetate (94–96). However, the use of anti-androgens as therapy may have undesired side effects.

In looking for transgenic animals of SBMA to perform our *in vivo* studies, all available SBMA transgenic mice with either truncated or full-length human AR (Table 1-4) were compared. Transgenic mice were first created using the full-length AR containing 45 CAGs, which is equivalent to the repeat length observed

in SBMA patients, driven by the interferon-inducible Mx promoter or the neuron-specific enolase (NSE) promoter. Expression of mutant AR was found in mice with the inducible Mx promoter, but at a lower level than normal endogenous expression. The mice demonstrated neither disease manifestations nor repeat length instability. Another transgenic mouse model was created with yeast artificial chromosomes (YACs) carrying the AR gene in the context of flanking non-coding sequences (82, 97).

Studies on independent lines of AR YAC transgenic mice carrying 45 CAGs, revealed intergenerational instability. However, this model failed to show the expression of mutant AR in RT-PCR or Western blot (WB) analysis. In order to enhance the toxicity of mutant AR, a transgenic mouse model was created with human AR containing 66 CAGs, which was longer than the longest repeat observed in SBMA patients, driven by the NSE promoter or the neurofilament light chain (NFL) promoter. Although expression levels of mutant AR were 2–5 times the endogenous AR levels, these mice showed no neurological symptoms, presumably because the CAG repeat was not long enough. Therefore, in 2002, Prof. Sobue laboratory generated transgenic mice expressing the full-length human AR containing 24 or 97 CAGs under the control of the cytomegalovirus enhancer and the chicken β -actin promoter (92). This model recapitulated not only the neurologic disorder, but also the phenotypic differences with gender, which is a specific feature of SBMA.

A full characterization of 97Q animals are summarized in figure 1-7 adapted from (92).

Reference	Transgene construct	CAG repeat instability	Motor impairment		Neuropathology		Muscle pathology
			Symptoms	Gender effect	Nuclear inclusions	Cell loss	
Bingham et al., 1995	Mx or NSE promoter, full-length human AR with 45 CAGs	(-)	(-)	(-)	(-)	(-)	(-)
LaSpada et al., 1998	YAC transgenic full-length human AR, with 45 CAGs	mild	(-)	(-)	(-)	(-)	(-)
Merry et al., 1996	NSE or NFL promoter, full-length human AR with 66 CAGs	(-)	(-)	(-)	(-)	(-)	(-)
Adachi et al., 2001	human AR promoter, sole 239 CAGs	mild	weakness, amyotrophy, incoordination	(-)	spinal cord, cerebrum	(-)	(-)
Abel et al., 2001	NFL promoter, truncated human AR with 112 CAGs	(-)	weakness, foot claspings	(-)	cerebellum	(-)	(-)
Abel et al., 2001	PrP promoter, truncated human AR with 112 CAGs	(-)	hypoactivity, foot claspings, tremor, seizure	(-)	spinal cord, cerebrum	(-)	(-)
Katsuno et al., 2002	chicken β -actin promoter, full-length human AR with 97 CAGs	(-)	weakness, amyotrophy	significant	spinal cord, cerebrum	(-)	grouped atrophy
McManamny et al., 2002	cytomegalovirus promoter, full-length human AR with 120 CAGs	(-)	weakness, amyotrophy, foot claspings	mild	brainstem	(-)	grouped atrophy fiber-type grouping hypertrophic fiber
Kennedy et al., 1968 Sobue et al., 1989	SBMA patients	mild	weakness, amyotrophy, fasciculation	significant	spinal cord, brainstem	(+)	grouped atrophy fiber-type grouping hypertrophic fiber

Table 1-4: Full characterization of introduced transgenic mouse models of SBMA (Adapted from (33)).

Table summarizing the list of introduced transgenic mouse models of SBMA. Length of the polyQ, and main clinical features of SBMA pathology in each line have been compared.

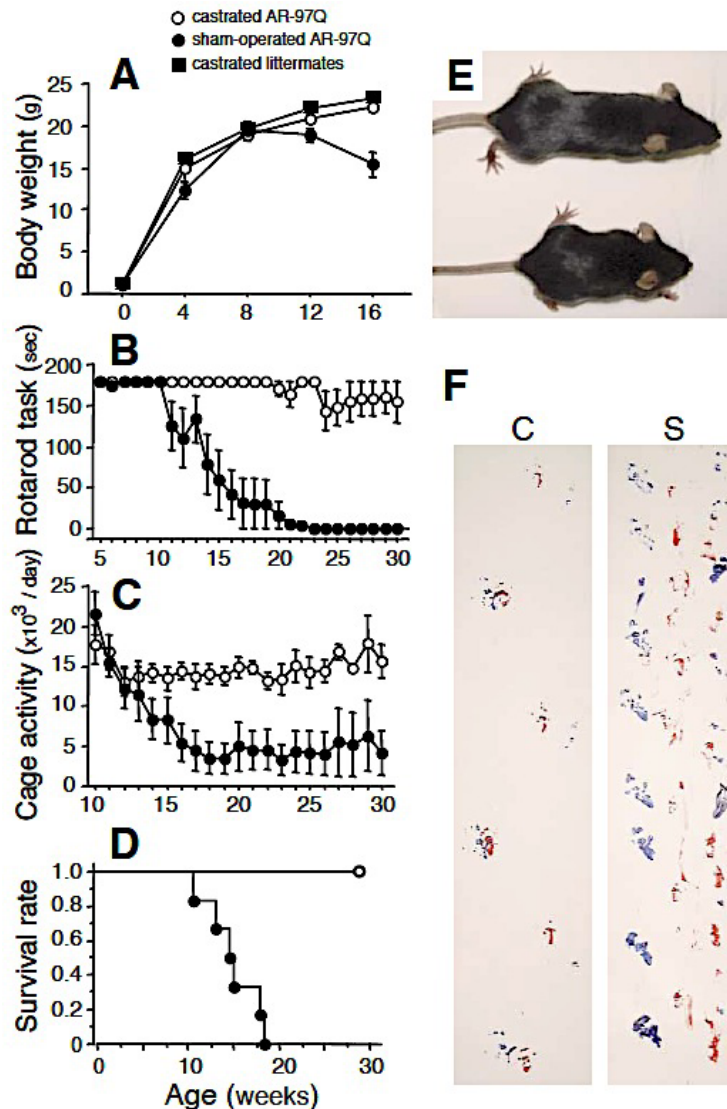


Figure 1-7: Effects of castration on symptomatic phenotypes of male AR-97Q mice. (A, B, C and D). Body weight (A), rotarod task (B), cage activity (C), and survival rate (D) of castrated (E, n = 6) and sham-operated (n = 6) male AR-97Q mice. All parameters are significantly different between sham-operated male AR-97Q mice and castrated male AR-97Q mice or castrated male littermates (n = 2) ($P = 0.0001$, $P < 0.0001$, $P = 0.006$, and $P = 0.0006$, respectively). (E) A castrated AR-97Q mouse (top) shows no muscular atrophy, which is striking in a sham-operated male AR-97Q mouse (bottom) (#2-6, 12-week-old). (F) Footprints of 12-week-old castrated (C) and sham-operated (S) male AR-97Q mice. Front paws are in red, and hind paws in blue. Adapted from (92).

Emerging evidence suggests a role for muscle in SBMA pathogenesis. Histological and molecular signs of muscle pathology are detectable before the appearance of pathological abnormalities in the spinal cord in a knock-in mouse model of SBMA (93), suggesting that mutant AR may exert a direct toxic effect on skeletal muscle.

In support of this notion is the observation that over-expression of normal AR in the skeletal muscle induces a phenotype similar to SBMA (84, 98). Analysis of muscle biopsy samples derived from SBMA patients suggests a mixed pathology with both myopathic and neurogenic features (10). However, the extent of weakness in later stages in SBMA is a consequence of motor neuron damage.

Knock-in mice expressing AR with 113 glutamines (AR113Q) develop early myopathy findings with little or no significant motor neuron loss until late in their disease course (93), consistent with muscle as a key site for SBMA disease pathogenesis. Moreover, while widespread transgenic expression of human AR20Q at levels comparable to endogenous AR does not produce a neuromuscular phenotype (82), skeletal muscle-specific overexpression of wild-type AR (AR22Q) in mice is sufficient to produce a SBMA-like neuromuscular disease, accompanied by denervation of target muscle and motor neuron axon degeneration, complete with androgen dependence, gender bias, axonopathy, and muscle wasting (84). Additionally, testosterone treatment of asymptomatic female transgenic mice overexpressing the AR22Q transgene in muscle, yielded pronounced neuromuscular deficits, but no detectable motor neuron pathology (83).

Although development of SBMA-like disease manifestations upon skeletal-muscle-specific expression of AR22Q may simply stem from the very high level of AR transgene overexpression in this model (84), the SBMA-like phenotype in males can be reversed upon cessation of testosterone treatment. These findings indicate that muscle-restricted AR toxicity may underlie SBMA disease pathogenesis and therapies targeting skeletal muscle may prove beneficial for patients. In support of this thesis, transgenic expression of anabolic insulin growth factor-1 (IGF-1) directed to muscle can rescue nerve pathology in SBMA transgenic mice, producing a significant extension in the life span (87).

1.2. Androgen receptor (AR)

1.2.1 AR function

The human AR is a 919-residue nuclear receptor (NR) that is expressed in most tissues including the brain, liver, kidneys, muscle, skin, bone and prostate and is important for the development of the male sexual phenotype (99, 100). Its gene is located in the long arm of the X chromosome, Xq11-12, and consists of 8 exons (Fig. 1-8). Exon 1 codes for the N-terminal domain (NTD), exons 2 and 3 for the DNA binding domain (DBD) and exons 4 to 8 for the hinge region and the ligand binding domain (LBD). In the late 1980s several groups cloned the human AR complementary DNA (cDNA), which paved the way for the structural and functional characterization of the protein (101, 102). Androgens mainly exert their

physiological roles via binding to the AR, which leads to nuclear translocation, binding to DNA and regulation of the transcription of AR target genes.

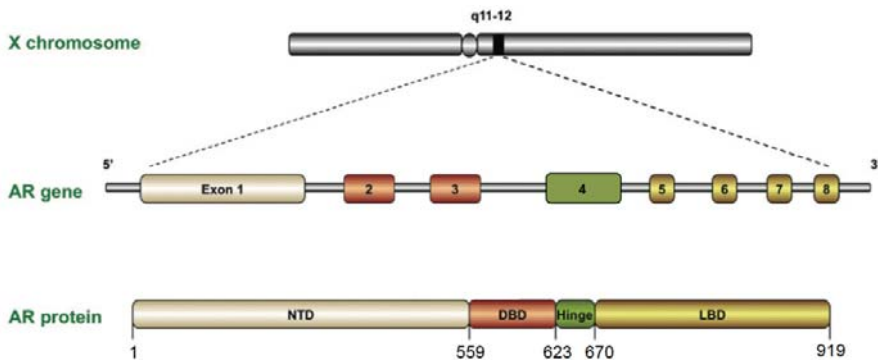


Figure 1-8: Schematic representation of the human AR gene and the AR protein Adapted from (103) .

1.2.2 AR structure

The AR is a member of the nuclear receptor (NR) superfamily, and belongs to the group of steroid receptors (SRs), together with the estrogen receptor (ER), the glucocorticoid receptor (GR), the mineralocorticoid receptor (MR) and the progesterone receptor (PR) (104, 105). Nuclear receptors are ligand-activated transcription factors and the members of this superfamily share a specific domain organization.

In AR, the NTD is located between residues 1 and 559 and contains activation function 1 (AF1), which is important for AR transactivation. The DBD between residues 560 and 622 contains two Zn fingers and mediates DNA-binding of the AR. A flexible hinge region (H, residues 623-670) separates the DBD from the LBD. Finally, the LBD is located between residues 671 and 919, is the domain of the protein to which androgens bind and contains activation function 2 (AF2) (Fig. 1-9).

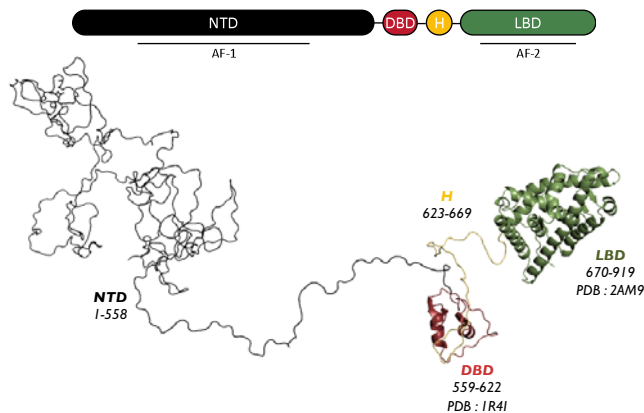


Figure 1-9: The domain structure of AR. a) The AR contains the NTD, the DBD, the flexible H, and the LBD. The structures of the folded domains resolved by X-ray crystallography and the corresponding residue numbers are indicated.

The main domain of AR involved in SBMA is the NTD, but a brief explanation about the interaction between C-terminal LBD and N-terminal domain FQNLF is important for a better understanding of the protective role of Hsp70, suggested by the results presented in this thesis.

Typically, the AF2 region of NRs interacts strongly with LxxLL signature motifs - in which L is leucine and X is any amino acid- present in several NR co-activators, like the family of p160 co-activators (SRC-1, SRC-2 and SRC-3) (106) As shown in figure 10, LxxLL motifs adopt a helical conformation and dock into the hydrophobic cleft formed by AF2 on the surface of the LBD. After hormone binding two charge clamp residues, K720 (in helix 3) and E897 (in helix 12), form part of the AF2 surface and stabilize further the interaction between AF2 and these co-activator motifs. The AF2 of AR is unique among NRs that it does not have a high affinity for such LxxLL signature motifs, even though it can accommodate LxxLL containing peptides. Instead, the AR AF2 preferentially binds FxxLF motifs, in which F is phenylalanine, L is leucine and X is any amino acid, as is the case in the ²³FQNLF²⁷ motif found in the NTD of the AR. This leads to an N/C interaction that is essential for AR transactivation and is regulated in a spatiotemporal fashion (107). This interaction will be discussed further in this section. FxxLF motifs are also present in AR co-regulators, such as ARA70, and can bind to the AF2 cleft (Fig. 1-10). The importance of these motifs in protein-protein interaction will be discussed further in results part.

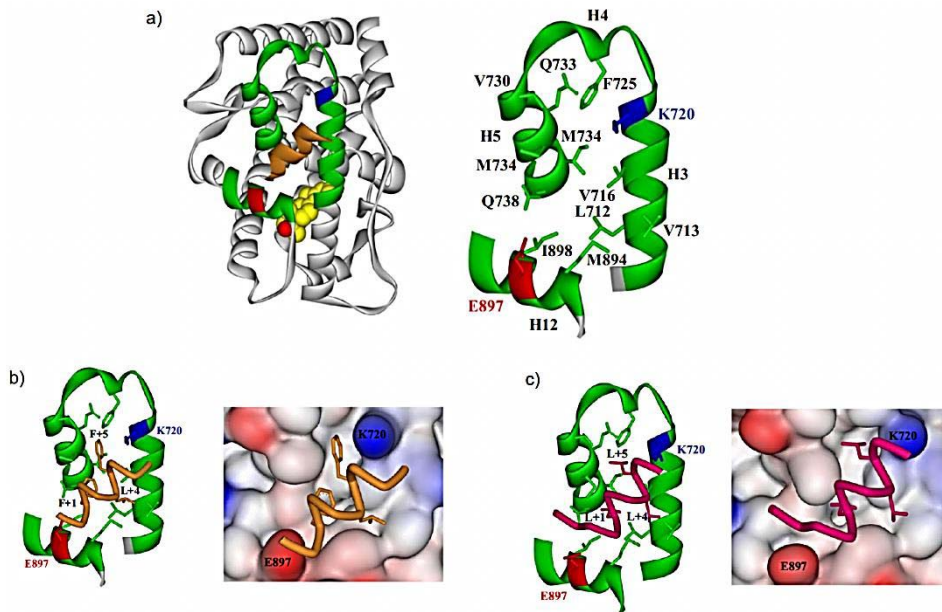


Figure 1-10: Structure of the AR FxxLF peptide bound to the AR LBD. (a) Global architecture of the AR FxxLF peptide (brown ribbon) and R1881 (space filled atoms; yellow, carbon; red, oxygen) bound to AR LBD (white ribbon) with helices 3, 4, 5 and 12 (green ribbon) forming the coactivator binding site. Conserved charged residues (blue, positively charged lysine; red, negatively charged glutamic acid) at the opposite ends of the groove are indicated. (b) Close-up view of the AR coactivator-binding groove. Helices 3, 4, 5 and 12 are shown (green ribbon). Side chains of residues composing the groove are shown as sticks. (c)–(f) Different representations of AR FxxLF (c and d) and TIF2 box III LxxLL (e and f) peptides bound to the AR LBD. Compared to FxxLF peptides, LxxLL peptides are shifted in the groove towards K720, resulting in different interactions, and thus stabilizations, between the two types of peptides with the AR LBD surface. Peptide backbones are shown as orange and magenta C-alpha coils, respectively. Side chains of the amino acid residues at positions +1, +4, and +5 are shown as sticks. The charged residues K720 and E897 are indicated for orientation. Adapted from (108).

In addition to these local effects, ligand binding also causes conformational changes in other domains of the AR. It exposes for instance the nuclear localization signal (NLS), located in part in the DBD and in part in the hinge region, which allows interaction with proteins that translocate the AR to the nucleus (104, 109). Ligand binding further induces the N/C interaction between the NTD and the LBD and the concomitant release of molecular chaperones, which changes the conformational properties of the NTD and possibly other parts of the AR (107, 108). Binding of the AR DBD to AREs does not only cause conformational changes in the DBD, but was also shown to induce a certain degree of helicity in the NTD (110).

This was proposed to act as a mechanism to regulate gene expression by enhancing the interaction of the NTD with co-activators (110–112). Furthermore, the AR NTD also modulates the DNA binding properties of the DBD by reducing its binding affinity for androgen response elements (AREs).

The ligand-induced N/C interaction between the NTD and LBD of the AR is the best characterized interdomain interactions of the protein. It has been described as necessary for its full transcriptional activity depending on the context, including the promoter sequence (113). It can occur by direct interaction between the NTD and the LBD or, instead, via the binding of SRC co-activators to both the NTD and the LBD of the AR. The regions of AR involved in the direct N/C interaction are the ²³FQNLF²⁷ motif in the NTD and the AF2 in the LBD.

The ²³FQNLF²⁷ motif is located close to the N-terminus of AR and plays an important role in transactivation. It is highly conserved among different species (114). Its binding to the AF2 is androgen-dependent and involves the folding-upon-binding of this region of sequence into a helix (115).

It was shown that amino acids 3-13 further modulated the N/C interaction without directly interacting with the LBD. The binding affinity between the ²³FQNLF²⁷ motif and the AF2 pocket was determined to be $1.2 \pm 0.2 \mu\text{M}$ by isothermal titration calorimetry (ITC) (116) and $9.2 \pm 0.4 \mu\text{M}$ by fluorescence polarization. Further binding partners of this specific sequence will be introduced in the result part.

Several amino acid repeats are present in the NTD, including two large polymorphic Gln and Gly repeats (residues 58-78 and residues 449-472, respectively), two shorter Gln repeats of six and five residues (residues 84-89 and residues 193-197) and short repeats of five Ala (residues 398-402) and eight Pro residues (residues 372- 379).

The roles of the polymorphic polyQ and polyG repeats have not been fully elucidated, but they have been proposed to provide inhibitory control over the NTD (117). Shorter polyQ and polyG repeat lengths are associated with increased AR activity, and deletion of the polyQ stretch results in a much more active AR. The N/C interaction was also found to be stronger for shorter repeat lengths. Furthermore, the length of both the polyQ and polyG stretch is inversely correlated with the risk of developing prostate cancer (PC) (118) and somatic mutations that shorten the CAG repeat are commonly observed in PC patients (119). In contrast, longer polyQ repeats correlate with an increased risk of SBMA and an earlier age of onset of this disease.

1.2.3 AR, a protein with a large intrinsically disordered domain

1.2.3.1 Intrinsically disordered proteins

Over the past century, evidence steadily was accumulated that a well-defined structure is the prerequisite of protein function. The success of the protein structure–function paradigm was thought to suggest that a protein could only function with a well-defined three-dimensional structure. This view is based on

more than 60,000 high-resolution structures in the Protein Data Bank, which in many cases enable the interpretation of function in terms of structure. Although deviations from this norm were always expected, they had, until recently, been neglected or ignored. There are many observations that argue for a novel family of proteins, which apparently exist and function without a well-defined structure. Based on predictions, we know that the structural disorder is abundant in all species, and due to its strong correlation with regulatory and signaling functions, its level is significantly higher in eukaryotes than prokaryotes (120). About 15-45% of eukaryotic proteins contain significant disorder of at least 30 residues in length (121, 122). Studies of these intrinsically disordered proteins (IDPs) are in the spotlight of current structural biology, leading to a rapid extension and transformation of the structure–function paradigm. These intrinsically disordered proteins/regions (intrinsically disordered protein IDPs/intrinsically disordered region IDRs) resemble the unfolded and denatured states of globular proteins. Unlike the latter, however, IDPs/IDRs carry out important functions often related to signaling or regulation (123) including specificity without strong binding, adaptability, involvement in post-translational modifications and rapid interaction with partners.

Therefore, the most important question of the field is the physiological functional mode of IDPs. Recent studies have indicated that structural disorder provides multiple functional advantages and IDP functions either originate from their disorder or from molecular recognition, when they undergo induced folding (disordered to ordered transition) upon binding to a partner protein (124, 125). These behaviors induce specific functional modalities, such as adaptability in binding, weak but specific binding and frequent regulation by post-translational modification. The concept of structure adaptability suggests that IDPs move to a structured state and become ordered upon binding to the partner. Binding by these motifs is usually weak, transient and possibly of limited specificity, which can become more specific and/or stronger, when influenced by their flanking regions (126–128).

The functional role of structural disorder from a biological process point of view addresses what type of cellular functions benefit most the lack of the stable structure. Several bioinformatics studies have indicated that IDPs are generally thought to be involved in processes of signaling and regulation as mentioned before.

1.2.4 Structural properties of AR-NTD

Unlike the DBD and LBD, the NTD has no stable secondary or tertiary structure. It is predicted to be an IDP and according to biophysical experiments it can be considered as IDP. The NTD of other NRs is similarly disordered and this flexibility is related to the function of this domain. The AR, in isolation, was shown to contain 13-16% helical secondary structure by circular dichroism (CD) and Fourier transform infrared spectroscopy (FTIR) (118, 129, 130). However, as described in

the previous section, interdomain communication and allosteric mechanisms can increase the helical content of AF1. The N/C interaction induces an α -helical conformation in the ²³FQNLF²⁷ binding motif and it has been proposed to affect the conformational properties of other regions in the NTD and possibly other domains of the AR. When the NTD is linked to the DBD, binding to DNA will also induce helicity in the NTD. This illustrates that in the context of full-length AR and/or when the AR is bound to DNA, the NTD adopts a more folded conformation than when it is studied in isolation. In addition, DNA binding of the AR and binding of co-regulatory proteins to the NTD can induce a more helical conformation and/or stabilize a particular conformation from this ensemble. This further facilitates interaction of other co-regulatory proteins and the assembly of a transcriptionally competent AR complex. Such a model, in which structure can be induced in the NTD and new binding sites can be created, allows for specificity and multiple target protein binding to the NTD. Some interactions possibly induce folding, whereas others require a more ordered conformation to interact with the NTD.

1.2.5 Importance of NMR to study IDPs

Although the initial name “unstructured” implied that IDPs might completely lack structure, it is now apparent that they have potentially function-related short and long-range structural organization that now can be studied in great detail by several experimental techniques with the most spectacular advances achieved through multi-dimensional NMR. NMR parameters that are sensitive to local structure, such as chemical shift, heteronuclear Overhauser effect (hetNOE), relation and residual dipolar coupling (RDC) values, can be determined. From a large number of possible disordered conformations, limited structural states are then selected to describe data collectively. IDPs are now also studied at the single-molecule level. Single-molecular fluorescence resonance energy transfer (smFRET) measurements, dSTORM or AFM also enables either the study of the dynamics of the structural ensemble of IDP molecules or the visualization of conformational changes. Therefore, we have used a selection of these approaches to study AR in SBMA pathology.

NMR plays a strategic role in the characterization of IDPs, as it is the only spectroscopic technique that can access atomic-resolution and structural as well as dynamic information on macromolecules. The key role of NMR in the characterization of the structure and dynamics of well-structured proteins has been discussed extensively in the last decades and the recent tremendous progress in this area is stimulating new experimental methods and computational approaches for the detailed characterization of IDPs/ IDRs and their interactions (131, 132). These atomic-resolution experimental results are going to contribute to the understanding of the functional role of dynamic and structurally heterogeneous protein modules, permitting the extension of the structure–function paradigm to also include the idea of disorder and flexibility. The first important consequence of the highly flexible nature of IDPs consists of extensive

conformational averaging, which drastically reduces nuclear chemical shift dispersion. Among the nuclear spins of interest for high-resolution protein NMR studies, ^1H is characterized by an intrinsically lower chemical shift dispersion, which, however, increases in ^{13}C and ^{15}N . The use of multinuclear experiments therefore had a great impact in protein studies, in general, and in the study of unfolded protein states, in particular, thanks to the improved chemical shift dispersion of heteronuclei compared to protons (133). More recently, the introduction of ^{13}C direct detection experiments permitted maximum advantage to be gained from the high heteronuclear chemical shift dispersion by only exploiting heteronuclear chemical shifts (133, 134). Once the sequence-specific assignment of the protein has been achieved, chemical shifts can be analyzed to understand whether specific regions of the polypeptide of an IDPs have secondary structure (135, 136). Chemical shifts, particularly heteronuclear shifts, do contain structural information and can be used as structural constraints to determine three-dimensional structures of folded proteins (137, 138). In the presence of extensive conformational averaging, deviations from random-coil chemical shifts can be used to estimate the so-called secondary structural propensities (SSP). Compact conformations typical of α -helices and the elongated conformations typical of β -strands do cause chemical shift variations in opposite directions for the various nuclei (139, 140). Therefore, when chemical shifts are available for several nuclei within each amino acid, they can provide information on the secondary structural propensities of different portions of the polypeptide, in particular when identified for several consecutive residues. This means that ^{13}C NMR can give information about chemical shifts of different nuclei and explores the presence of motifs with secondary structure propensities in sequences predicted as ID.

1.2.6 Use of NMR in the characterization of IDPs

Once the protein has been expressed and purified as well as enriched with stable isotopes (^{15}N and ^{13}C), the next step consists of evaluating the feasibility of a complete structural and dynamic characterization through NMR. Then planning the proper set of experiments to perform a complete sequence specific assignment of the protein and to determine additional observables that can provide structural and dynamic information takes place. The most suitable NMR experiments to evaluate the feasibility of a complete NMR characterization of the protein are the two-dimensional experiments correlating the amide nitrogen with the directly bound amide proton (^1H - ^{15}N heteronuclear single quantum coherence- HSQC).

As extensive overlap may occur in IDPs, it might also be interesting to look for cross-peaks of residues with characteristic chemical shifts falling in isolated regions, such as those of Gly residues in HN HSQC/CON spectra and of Pro residues in CON spectra. The number of observed versus expected cross-peaks already gives an indication on whether it will be possible to characterize the entire protein. From a more technical point of view, it is important to evaluate the

resolution and overall time to acquire a two-dimensional experiment as well as to evaluate the best experimental approach and pulse sequence to identify a specific kind of correlation. As an example, heteronuclear correlations involving backbone nitrogen nuclei, both for HN and CON, can be detected with several different variants of the experiments that may help to increase the sensitivity, increase the resolution, decrease experimental time and increase the number of detected cross-peaks. Provided the two-dimensional experiments can be collected with good sensitivity and resolution in a limited amount of time and that a large share of the expected cross-peaks can be identified, complete structural and dynamic characterization of the protein is feasible. The spectral quality can often also be improved through minor changes in the experimental conditions (temperature, pH, buffer, salt, etc.), of course remaining in a meaningful range. This therefore constitutes another important check to be performed before collecting the NMR experiments. Finally, determination of the ^{15}N relaxation rates provides accurate information on the motional properties of the backbone for each amino acid relatively quickly. The ^{15}N relaxation rates are also very useful to have a general estimate of the expected transverse relaxation rates, which have a large impact on the overall sensitivity of multidimensional NMR experiments.

The suite of experiments generally used for the sequence-specific assignment of well-folded proteins can also be applied to disordered ones, taking care to optimize the experimental setup for the resolution (i.e., a high number of acquired data points). The favorable relaxation properties of IDPs and IDRs generally provide good sensitivity, enabling the optimization of the experiments in terms of the resolution (i.e., compromising the number of scans for the number of data points). The list of the most suitable three-dimensional NMR experiments either based on ^1H detection or on ^{13}C detection, as well as the correlations expected in each experiment have been reported (141) The necessary number of experiments of course depends on the complexity of the protein. For a protein of intermediate size, a subset of those indicated may be sufficient, while for larger proteins the more information that can be collected, the more robust will be the resulting sequence-specific assignment. Counter-intuitively, a larger set of three-dimensional spectra to be recorded and analyzed (which at first glance might seem to imply more experimental work) actually reduces the time necessary for the assignment, as there are more chances to solve ambiguities. Recently, several approaches to reduce experimental time by decreasing the amount of acquired data points in indirect dimensions of multidimensional NMR experiments or the combined evolution of different shifts in one dimension have actually enabled the collection of multidimensional NMR experiments with higher dimensionality than three-dimensional in a reasonable amount of time. These experiments have been proposed both in the ^1H - and ^{13}C -detected modes and will have an impact on reducing ambiguities and in speeding up the assignment procedure. In addition to chemical shifts and ^{15}N relaxation rates that we used in this thesis, a variety of different observables can be determined that contain information on different features of the environment experienced by different nuclear spins. While a lot of

information is directly available from chemical shifts and relaxation rates, a very accurate picture describing the structural and dynamic properties of the conformational ensemble can be achieved by a comparative analysis of additional NMR data, such as exchange effects with the solvent, scalar couplings (both 3J and, if measurable, scalar couplings mediated by hydrogen bonds), residual dipolar couplings, 1H - 1H NOEs, and PREs. The combined qualitative interpretation of this set of experimental data is able to give a very good idea of the structural and dynamic properties of the protein. As handling structural and dynamic aspects in a quantitative way is not trivial, several elegant approaches to also obtain quantitative information have recently been proposed. These either enable the selection of a subset from a large ensemble of generated conformers, which can reproduce experimental data, or make use of molecular dynamics simulations to account for flexibility and conformational changes. Due to the complex NMR experiments and limitations of time, we limited our approaches to use chemical shifts and relaxation rate in this work, and, fortunately, we have obtained a complete dataset of all necessary information for dynamic and structural characterization of AR N-terminal domain constructs with either 25 or 4 Gln residues.

1.3. PolyQ proteins and coiled-coil interactions

PolyQ proteins belong to protein families with many different functions, but a high number of them are transcription factors and nuclear proteins (Fig. 1-11 and 1-12) (142). The conservation of polyQ regions and their evolutionary patterns in protein families such as Htt and others, suggest that they have an important biological function (Fig. 1-14). Encouraged by this evidence, the function of polyQ tracts by association to features in proteins bearing polyQ tracts or interacting with those, have been recently studied (45). And the following strong correlations were found: (i) proteins with longer polyQ tracts have a higher number of interactors than proteins with short polyQ tracts or without such sequences and (ii) there is a statistically significant presence of predicted coiled-coil (CC) regions next to polyQ regions. A CC domain consists of parallel or antiparallel α -helices that further fold like the strands of a rope (Fig. 1-12). These domains facilitate protein dimerization and are present in proteins with many biological functions (143).

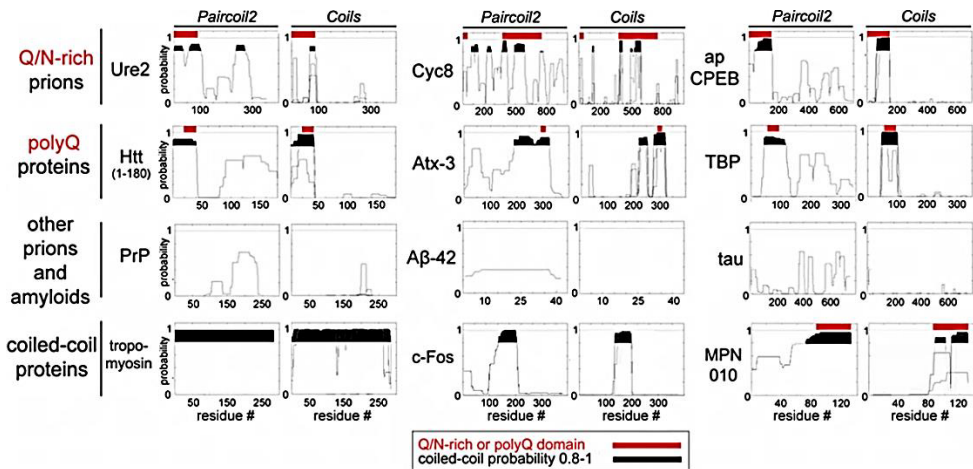


Figure 1-11: Overrepresentation of CCs in Q/N-rich and PolyQ Proteins and Their Interactomes (A) Per-residue CC probability (0 to 1) for Hsp104, CHIP, PQBP, and HIP-1 obtained with Coils. Peaks with the highest probability (0.8–1) are highlighted in black. (B) Proportion of CC proteins among interactors of Ure2, apCPEB, and Htt, and among known and candidate yeast prions (144), and human polyQ expansion proteins, as compared with eukaryotic proteomes. (C) CC probability for prions and amyloidogenic proteins, compared with proteins with known CC structure, obtained with Paircoil2 and Coils (145). Red bars represent Q/N-rich (>15%) or polyQ regions.

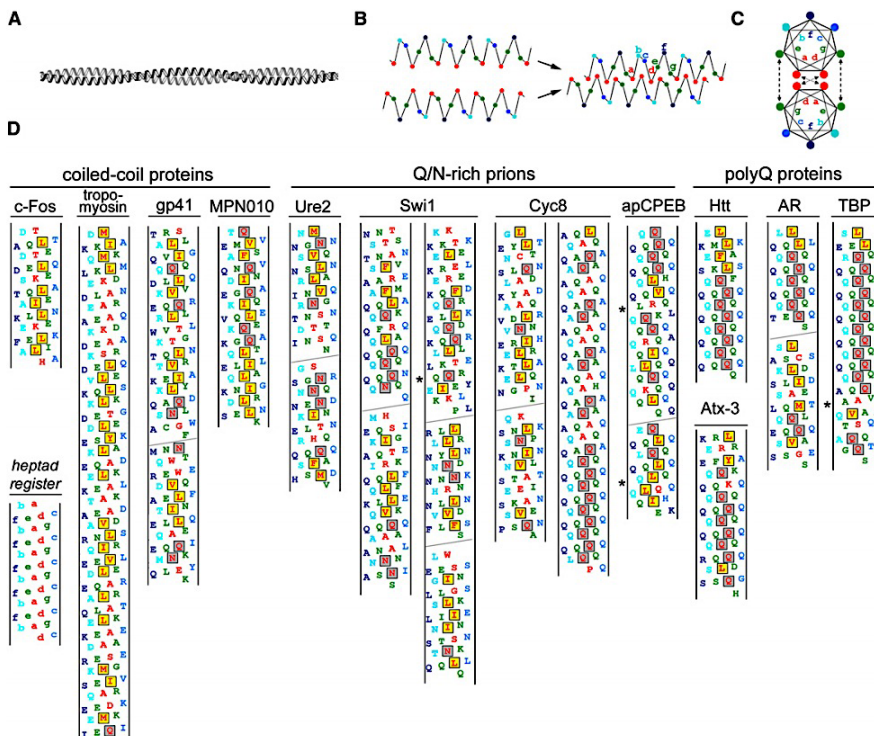


Figure 1-12: Heptad Repeats in Q/N-rich and PolyQ Proteins (A) Scheme of two coiled alpha-helices. (B and C) Lateral and zenithal view of two coiled helices. Red circles, heptad positions a/d; green circles, g/e; and cyan-to-blue circles, b, c, and f. (D). Note the vertical alignment of a/d residues, and the grouping of a/d hydrophobic residues in discrete clusters along the helices (146).

Coiled coils (CCs) are α -helical super secondary structures mediating protein-protein interactions (Fig. 1-13), oligomerization, and other functions through the coiling of helices belonging to the same or different polypeptide chains (146). The finding of CC proteins in the interactomes prompted the researchers to look for CC regions in the Q/N-rich and the polyQ proteins themselves, which could serve both as substrates of interaction with other CCs and mediators of aggregation. Previous analysis revealed heptad repeats typical of CCs in regions flanking or overlapping with Q/N-rich domains and polyQ tracts (Fig. 1-12) (45–47).

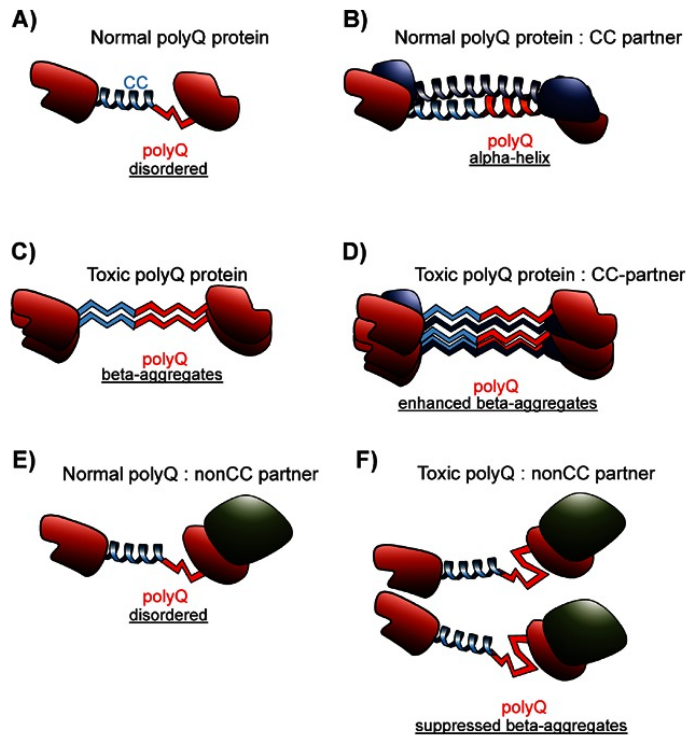


Figure 1-13. Cartoon models of interactions of polyQ Wt and mutant proteins with CC and non-CC partners. (A) A wild-type polyQ protein contains a CC region near a disordered polyQ tract (red). (B) Upon interaction with a CC partner (blue), the polyQ region adopts a CC conformation and modulates the interaction. (C) The mutant polyQ protein with an expanded polyQ tract adopts a β -strand conformation and forms β -sheet aggregates. (D) The CC partner of the wild-type polyQ protein interacts with the aggregates and enhances them. (E) A non-CC partner of the wild-type polyQ protein (green) binds at a different domain. (F) This non-CC partner binds to the aggregates of the mutant polyQ protein, sequesters the toxic protein and sterically impedes its aggregation. Adapted from (47).

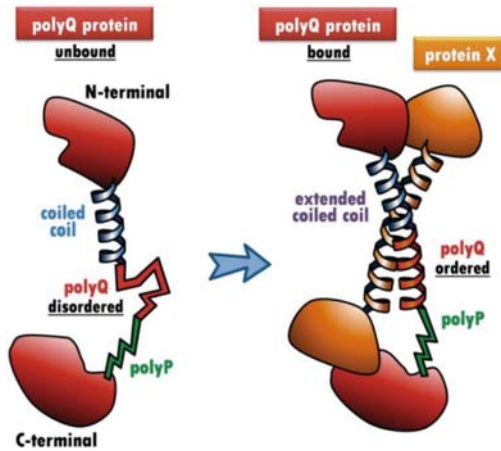


Figure 1-14. Cartoon of proposed polyQ function in protein-protein interactions. Left: a polyQ protein contains a coiled-coil (blue), followed by a polyQ region (red) and a polyP region (green). In the unbound state, the polyQ region is disordered. Right: upon interaction with a protein partner X, the polyQ region adopts a coiled-coil structure that extends the original coiled-coil. The polyP region remains unstructured capping precisely the extension of the coiled-coil. Adapted from (45).

In conclusion, several lines of evidence indicate that: (i) CC proteins that interact with a wild-type polyQ protein also interact with the corresponding pathogenic polyQ-extended protein, and (ii) these CC protein partners promote the aggregation of the pathogenic polyQ-extended protein. A careful analysis of these interactions should help us to develop therapeutic strategies that might prevent neurodegeneration in humans. More information on role of CC super secondary structure will be discussed in results.

1.4. Introduction to heat shock proteins and to their involvement in protein misfolding diseases

1.4.1 Protein quality control

The three dimensional fold of proteins is determined by the protein's amino acid sequence (147). However, since the cellular environment is extremely crowded, an unassisted folding is often not feasible and could lead to potentially cytotoxic protein aggregation (148). Moreover, throughout their lifetime proteins may become damaged by various stresses (149). Therefore, a sophisticated protein quality control (PQC) system has evolved to assist folding of proteins and prevent cellular dysfunction caused by damaged proteins and other aberrant proteins resulting from errors in transcription, translation or which become misfolded due to the presence of genetic mutations, as observed in most of amyloid diseases.

1.4.1.1 Molecular chaperones

The molecular chaperones represent a first line of defense in the PQC and have assisted in the folding of polypeptides from the moment they are translated by the ribosomes (148). They bind to the folding intermediates of polypeptides, preventing their aggregation and in some cases, actively assisting them in the folding through cycles of binding and release as in Hsp70 machinery system (148,

150), but do not become part of the final functional structure of protein complexes (151). If the native conformation of the proteins cannot be achieved, chaperones can facilitate degradation, either passively by preventing aggregation, and keeping their clients competent for proteolysis or actively by directing clients to degradation machines (150).

1.4.1.2 Hsp's and their function

The heat shock proteins (Hsp) are a family of structurally unrelated proteins, belonging to the much larger family of the molecular chaperones. The human family comprises 5 subgroups, including the HspA (Hsp70), HspB (small Hsp), HspC (Hsp90), HspD (Hsp60), HspH (Hsp110) and the DNAJ (Hsp40) families of proteins (152). The number of genes encoding the Hsp family varies largely for each organism, ranging for the DNAJ proteins from 6 in *E.Coli* to 50 in *H. Sapiens* (153). It is not clear why the genome encodes such a large amount of Hsp members, but compartmentalization, cellular specialization (substrate specificity) and developmental (regulation) purposes may explain such an expansion (153). In addition, evolution may have occurred for functional diversity (substrate processing). Hsps have been named after their inducibility upon heat shock, and this is how they were initially defined (154). But as the human genome was uncovered, it was found out that many additional sequence or structural homologs existed, among those, some are not heat inducible at all. Nowadays, we know that many members are also constitutively expressed in different cell compartments, induced by many different forms of either intrinsic or external forms of stress or up-regulated under specific conditions of differentiation (153).

1.4.1.3 The Hsp70 machine

The heat inducible Hsp70 (HspA family) and the constitutively expressed cognate HSC70 are involved in many cellular processes affecting protein folding and unfolding (155). The Hsp70's do not work alone, but work together with several chaperones, co-chaperones and cofactors forming an 'Hsp70 machine' (150). So far, all characterized Hsp70's can, in a rather non-selective and promiscuous manner, bind substrates through their C-terminal substrate binding domain (SBD) and prevent aggregation of the substrates by providing a time window for the client proteins to obtain their native conformation. The binding/release cycle of Hsp70 family members is regulated by nucleotides. When ATP is bound (to the N-terminal ATPase domain) Hsp70s are in an "open conformation" with low substrate affinity allowing fast substrate binding and release (Fig. 1-15). Upon ATP hydrolysis, which is stimulated by both client binding and by DNAJ proteins (Hsp40 family), a conformational change in the peptide binding domain is induced that closes the so-called "lid" and results in a high affinity ADP-bound 'closed conformation' stabilizing substrate binding (148, 156, 157) Upon ADP-ATP exchange, which is regulated by nucleotide exchange factors (NEF), the lid opens up again and substrates can be released. If not yet folded properly, clients may rebind and the cycle starts again until functional folding occurs or, if folding

remains unsuccessful, released clients are recognized by the proteolytic systems and subsequently degraded (150).

It is still not well understood, whether all Hsp70 family members have the same mode of action, are functionally interchangeable or use all the same co-factors (153). Except for two of family members (HspA12 and HspA14), which lack the conserved substrate binding domain, the sequence similarity of all human Hsp70s is very high and all of them are highly homologous to Hsp70 from other species, including *E.Coli*, suggesting limited functional differentiation (150).

1.4.1.4 DNAJ proteins (Hsp40s): the co-regulators of the Hsp70 machinery system

Apart from the presence of the J- domain that identifies the DNAJ proteins, in contrast to Hsp70 family, the DNAJ family is more heterogeneous, with various additional domains (150). The DNAJ proteins were primarily divided into three groups: DNAJA, DNAJB and the DNAJC proteins, but this classification does not relate to the biochemical function. With respect to their function, the DNAJ proteins can be divided into four groups: DNAJ proteins with promiscuous-, selective, no-substrate binding and a small group with unclear capacity to bind substrates (150). Within the group of promiscuous binders, some DNAJ proteins have an important role in de novo protein folding (158, 159), while other members have specific functions in preventing protein aggregation (38) and/or support protein degradation(160) or specifically target substrates towards degradative pathways (161). Therefore, the DNAJ proteins are considered to be the responsible entities for the multi-functionality of the Hsp70 machine (150). Figure 1-15, models the co-regulatory function of Hsp40 and Hsp70 in protein folding process.

Among DNAJ family, previous studies have indicated that DNAJB1 play an important role in aggregation suppression and toxicity of polyQ proteins. DNAJB1 suppressed AR aggregation and toxicity in cell models, probably by enhancing the degradation of soluble AR (162, 163). DNAJB1 was also protective against Htt aggregates and toxicity in various cellular models (40, 164, 165) Clearly, the DNAJB1 action was fully dependent on its interaction with Hsp70 members, as J-domain mutant without full functional collaboration with Hsp70 were ineffective (38). The protective effect of DNAJB1 against Htt aggregation was also lost when longer polyQ stretches were used (38). Yet, DNAJB1 could also partially suppress neurodegeneration in HD flies (166). Overall, DNAJB1 seems a rather generic suppressor of polyQ aggregation and toxicity. However, dependent on interaction with Hsp70s, its action generally does not require Hsp70 co-upregulation. Given the fact that DNAJB1 is the most abundantly expressed DNAJB member, including the brain (165), it may represent a first line of defense against polyQ pathology, which fails in the course of time or when the folding problems exceed its capacity (longer Q repeats). Not only Hsp40, but also nucleotide exchange factors play an important role in the functionality of Hsp70.

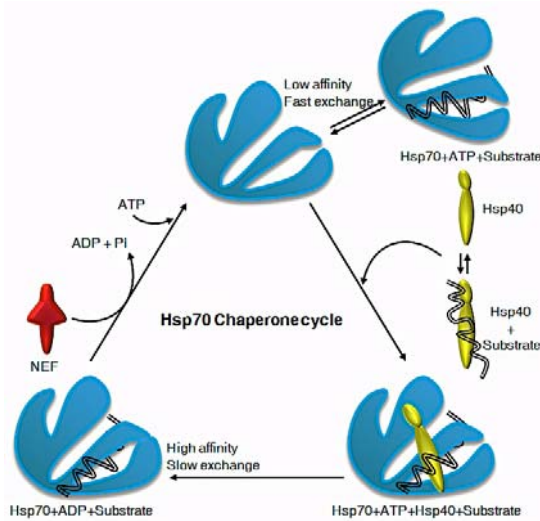


Figure 1-15: Hsp40 (with a conserved J-domain) interacts with Hsp70. During protein folding process Hsp70s do not function individually. They form a complex with J-proteins to form the folding-machine. Since the intrinsic ATPase activity of Hsp70 is very weak for stoichiometric coupling to the substrate binding-release cycle, J-proteins act as accessory factors for Hsp70 and can serve more than one Hsp70 molecule. They stimulate Hsp70 ATPase activity by transiently interacting with it. In a chaperone cycle, Hsp70 proteins have two distinct conformations; ATP-bound state that transiently interacts with client proteins and ADP-state that stably bind to the substrate. A client protein interacts with Hsp70 in its ATP bound state. Few subclasses of J-proteins directly sequester substrates into the peptide-binding cleft of an Hsp70 and are coupled to the stimulation of ATP hydrolysis by J-domains of Hsp40s that interact with the ATPase domain of Hsp70. The conversion of Hsp70 to ADP bound state stabilizes the interaction of Hsp70 with the client protein. Nucleotide exchange factors replace ADP with ATP resulting in dissociation of the bound peptide and hence prime Hsp70 for a second cycle of interaction. Adapted from Hsp information resource (156).

1.4.1.5 Nucleotide exchange factors (NEF's): the co-regulators of the Hsp70 machine

Despite the fact that the DNAJ proteins may be the main drivers of the functional diversity of the Hsp70 machines, the NEFs may also play important roles. In humans, 4 different groups of NEF's exist working via different mechanisms of nucleotide exchange (167). These 4 different NEF's are HspB1 and BAP, the HspA like NEFs, consisting of HspH1 - HSPH4 (Hsp110), and the BAG family (Bcl-2 associated athanogene, BAG1-6). In contrast to the other NEF's, HSPB1 and BAP only seem to function in protein folding (150, 168–170). The BAG family members 1, 2, 3 and 6 interact directly with HspA (Hsp70) (168). BAG-1 is the best-studied member and acts, depending on its concentration, as a positive (low levels) or negative (high levels) regulator of folding reactions. It was also shown that BAG-1 interacts with the proteasome and act as a coupling factor between the folding and the proteasomal system (171). Another BAG family member (BAG-3) has been linked to the stimulation of autophagy, although this seems linked more to its association with small Hsp rather than with Hsp70s (172). How other members of the BAG family affect Hsp70 dependent reactions remains to be elucidated.

1.4.2 Protein folding diseases

The long-term health of the cell is dependent on a proper function of the protein quality control (PQC) system, generating a protein homeostasis where biosynthesis, folding, translocation, assembly/disassembly and clearance are tightly balanced (173). Aging is accompanied by a decline in the cellular proteolytic capacity, which causes defects in the chaperone induction, function and capacity (174). In parallel, the protein damage accumulating during ageing, results in an overload of the cellular PQC (174), which is even more pronounced in post-mitotic cells, such as neurons and muscles. This overload of neuronal PQC during aging may be one of the reasons for age related onset of several misfoldings related to neurodegenerative disorders like Alzheimer's disease and Parkinson's disease. Moreover, genetic disorders like polyQ diseases are also characterized by an age-related onset, usually earlier than non-hereditary folding diseases due to an earlier occurrence of the protein homeostasis imbalance caused by the expression of the diseases protein chronically (165).

1.4.3 Heat shock proteins and polyQ degradation

The causative mutations in AR that results in expansion of the CAG repeat and which promotes hormone dependent AR misfolding and oligomerization are critical to toxicity. In parallel to our observations from other studies, we have focused on understanding the proximal mechanism that regulates the degradation of polyQ AR, with the goal of finding a pathway to diminish amounts of toxic protein and ameliorate the disease phenotype. The nuclear inclusions consisting of the mutant AR protein are characteristic and combine with many components of ubiquitin–proteasome and molecular chaperone pathways, raising the possibility that misfolding and altered degradation of mutant AR may be involved in the pathogenesis. It has been reported that the overexpression of Hsp chaperones reduces mutant AR aggregation and cell death in a neuronal cell model (175, 176). To determine whether increasing the expression level of chaperone improves the phenotype in a mouse model, the authors cross-bred SBMA transgenic mice with mice in Prof. Sobue's laboratory in Nagoya overexpressing the inducible form of human Hsp70 and demonstrated that high expression of Hsp70 markedly ameliorated the motor function of the SBMA model mice (86, 163). In double-transgenic mice, the nuclear-localized mutant AR protein, particularly the one of the large complex form, was significantly reduced. Monomeric mutant AR was also reduced in amount by Hsp70 overexpression, suggesting the enhanced degradation of mutant AR. These findings suggest that Hsp70 overexpression ameliorates the SBMA phenotypes in mice by reducing nuclear-localized mutant AR, probably caused by enhanced mutant AR degradation (163). This study provided the basis for the development of an Hsp70-related therapy for SBMA and other polyQ diseases.

Another main cellular machinery with the dominant role of regulating proteostasis of the AR is the Hsp90 and Hsp70- based chaperones which function together as a

There are many well-studied works published by different groups, nevertheless, there are still some controversies in the field, which makes it difficult to look for some specific therapeutics strategies dealing with Hsps and their interaction with AR. Therefore, we proposed high resolution data on interactions of AR with different Hsps to answer many of yet unsolved controversies but also in order to use this approach for lead compound studies and drug design in SBMA and other polyQ diseases.



PART II.

A I M A N D O B J E C T I V E S

Chapter 2. Thesis aim and objectives

2.1. Thesis objectives

Chapter 2. Thesis aim and objectives

2.1. Thesis objectives

A multi-disciplinary approach combining structural Biology, Biochemistry and Biomedicine was used in the experimental part of this work. To achieve our aim we have divided the work in 2 main objectives.

The goal of this thesis is to study the structural aspects of AR aggregation *in vitro* and in a transgenic mouse model of SBMA. For the first goal, we expressed and purified two constructs of the intrinsically disordered N-terminal domain of AR of approximately 150 amino acids comprised of either 4Q (135aa) or 25Q (156aa). We carried out a thorough high-resolution characterization of both constructs using NMR. Furthermore, we mapped the binding epitopes of interaction between 2 main Hsps involved in the polyQ disease, Hsp72 and Hsp40, using NMR. For the second aim, we used human AR expressing mouse lines (97Q-24Q) to investigate the aggregates formation of AR and characterize the structural properties of aggregates isolated from transgenic mouse tissue homogenates (Including mainly spinal cord and muscles). 97Q transgenes with longer polyQ show dramatic movement impairment during their life span, while the 24Q doesn't show any phenotype; therefore this model is the most used transgenic model among different research groups.

Moreover, similar to the human patient cases, AR intra-nuclear aggregates can be detected using various specific anti- AR antibodies.

First aim: Structural characterization of 4Q and 25Q by NMR and mapping of binding epitopes of Hsp72 and Hsp40

We were primarily interested in finding out the role of the flanking regions of polyQ in AR as in the case of Huntington disease; they have shown to play an important role in aggregation properties of N17 construct of Htt. To study the initiation of the aggregation formation of AR and the role of flanking regions, we have expressed a construct of N-terminal part of AR that comprises either 4Q or 25Q. The construct with 4Q gives us the opportunity to study the flanking regions of protein without bringing the protein to the aggregated state. We expressed ¹⁵N, ¹³C labeled protein; purified and assigned all the nuclei (more than 97%) and by using the chemical shift information we have managed to find the secondary structure propensities of 4Q. We also have expressed, purified and surprisingly managed to assign the larger construct with 25Q and characterize the secondary structure properties of this construct. Comparing 4Q with 25Q, we indicated the effect of increase in the number of the Q in the secondary structure propensities. SSP (secondary structure propensities) plot from 4Q show an alpha-helical region

in the ⁵⁵LLLL⁵⁸ region just flanked to 4Q, while increasing the number of polyQ from 4 to 25 led to a dramatic increase in the alpha helical propensity of 25Q, which initiates from ⁵⁵LLLL⁵⁸ and continues along the polyQ region. Therefore, we can conclude that ⁵⁵LLLL⁵⁸ motif is the initiation of α -helix formation in a 4Q construct, while increasing the number of polyQ in the context of 25Q construct stabilized the α -helix in the polyQ region. These results are the first high-resolution report of the full backbone assignment and structure of a polyQ protein from the group of 9 polyQ proteins. While the role of the polyQ stretch in normal AR activity remains unknown, expansion of the polyQ stretch clearly have profound effects on the function of AR. And therefore, we have also compared the relaxation rates of N-terminal domain construct of AR in 4Q and 25Q by using hetero-nuclear NOEs, R1 and R2. The R2 data indicate 3 main regions of interest with less flexibility in the constructs, including ²⁴FQNLF²⁸, ⁵⁵LLLL⁵⁸ and ⁸⁵GPTGYL⁹⁰. Among the 3 regions, ⁵⁵LLLL⁵⁸ region followed by polyQ in 25Q shows a dramatic increase in rigidity when the number of Q increases.

Having obtained the backbone assignment of these proteins, we mapped the binding epitopes of Hsp72 and Hsp40 on AR- 4Q and AR-25Q. ²⁴FQNLF²⁸ and its close flanking regions is the binding motif shared between Hsp72 and Hsp40, while Hsp40 additionally binds to ⁵⁵LLLL⁵⁸ and polyQ regions and work in close collaboration with Hsp72.

Second aim: Structural characterizing of the tissue homogenate from 97Q mice, comparing with 24Q and Wt mice, quantifying the aggregate loads of 97Q and find the toxic species among different tissue homogenates.

For this purpose we have taken tissue homogenate from 97Q, 24Q and Wt mice and used our biochemical and biophysical methods to find the fraction of tissue homogenate that can play the critical role as a seed (the nucleus for aggregation formation) for intranuclear AR aggregates. The challenge was mainly to prepare the tissue homogenate and detect the aggregates. We tried different methods and experimental conditions, which finally succeeded to detection of not only AR aggregates in the stacking part of WB gels (due to their SDS resistance nature) but also the monomeric form of the 97Q and 24Q AR.

We also have quantified the aggregates in different tissues using Seprion ELISA kit. Seprion ELISA kit works very similarly to the conventional ELISA except for the presence of a specific hydrophobic Seprion ligand at the surface of the 96 well microplates and selectively recognizes the aggregated form of the protein based on the arrangement of polar and hydrophobic groups in aggregates, that will not occur in monomeric form of the protein. By using the Seprion ELISA, we have shown that there is a significant difference between the tissue homogenates of 97Q and 24Q animals. With this optimized method, we are now able to quantify the aggregation amounts in different tissues and this would certainly be a useful read out for a drug/lead compound versus placebo study.

We used similar method but with magnetic bead system (PAD beads) and isolated the aggregates from spinal cord and muscle tissues of 97Q mice in three different groups of young (before the onset of the disease), middle age (at the beginning of the onset of the disease) and old (without any cage activity). We used atomic force microscopy and investigated the aggregated species in different tissues. Surprisingly, we found two totally different species in spinal cord and muscle, and additionally we have found a correlation between the age of the animals and the length and number of aggregates in 97Q muscle and spinal cord. Now we have a clear idea about the structure of these species, their size and even about their toxic effects in Neuroblastoma cells.

PART III.

MATERIALS AND METHODS

Chapter 3. Materials and Methods

- 3.1. Molecular Biology methods
- 3.2. Protein expression and purification
- 3.3. Isothermal Titration Calorimetry (ITC)
- 3.4. Fluorescence Polarization Substrate Competition (FPSC)
- 3.5. NMR spectroscopy of biomolecules
- 3.6. Transgenic mouse colony maintenance
- 3.7. Biochemical characterization of tissues
- 3.8. Quantification of aggregates in tissue
- 3.9. Isolation of aggregates species from *ex vivo* material
- 3.10. Cell culture and Immuno-Fluorescence staining
- 3.11. High-resolution microscopy of aggregated species
- 3.12. Atomic Force Microscopy (AFM)
- 3.13. Transmission Electron Microscopy (TEM)
- 3.14. Immuno-histochemical analysis
- 3.15. Synchrotron micro-computed tomography
- 3.16. Statistical analysis

Chapter 3. Materials and Methods

3.1. Molecular Biology methods

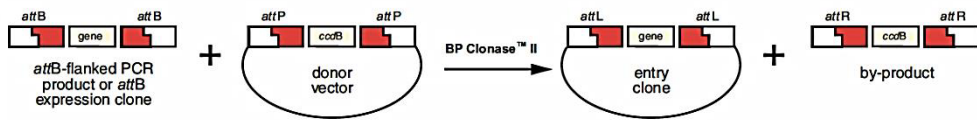
3.1.1 Recombinant Gateway technology

The recombinant Gateway® Technology with Clonase™ II of LifeTechnologies (Cat.No. 12535-029 and 12535-037) was used for cloning of 2 AR constructs, named 25Q and 4Q with an N-terminal HisMBP-tag. This technology allows cloning of DNA sequences of interest via recombination into a donor vector and subsequently transferring it from the donor vector to a wide range of available destination vectors (bacterial expression vectors, mammalian vectors and Baculovirus vectors).

The Gateway technology is based on the bacteriophage lambda site-specific recombination system that facilitates the integration of lambda into the *E. coli* chromosome. The recombination involves two major components: the DNA recombination sequences (attachment sites (att sites)) and the enzymes that mediate the recombination reaction, i.e. Clonase II enzyme mix.

The lambda recombination occurs between site-specific att sites: attB sites on the *E. coli* chromosome and attP sites on the lambda chromosome. The att sites serve as the binding site for the recombination enzymes. These enzymes bring together the target sites to which they bind, cleave them and covalently attach the DNA again after the recombination. The DNA segments flanking the recombination sites are switched such that, after recombination, the att sites are hybrid sequences comprised of sequences donated by each parental vector. For example, a recombination between attB and attP sites gives rise to attL and attR sites. The two recombination reactions that constitute the basis of the Gateway technology, the BP reaction and the LR reaction, are based on this recombination principle (Fig. 1). The BP reaction facilitates recombination of an attB substrate (attB-PCR product or a linearized attB expression clone) with an attP substrate (donor vector) to create an attL-containing entry clone (Fig. 3-1a). A BP Clonase II enzyme mix catalyzes this reaction. The LR reaction facilitates the recombination of an attL substrate (entry clone) with an attR substrate (destination vector) to create an attB-containing expression clone

a) BP reaction



b) LR reaction

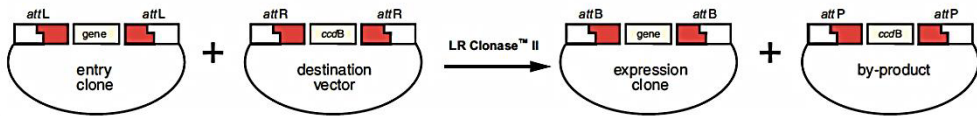


Figure 3-1: The Gateway recombination reactions. a) The BP reaction inserts the gene of interest, flanked by attB recombination sites, in a donor vector that contains attP recombination sites. This creates an entry clone containing the gene of interest. b) The LR reaction to further transfers the gene of interest from the entry clone (attL recombination sites) to a destination vector (attR recombination sites) to create an expression clone containing the gene of interest. Figure adapted from the Gateway manual available on the website of Life Technologies.

The pDONR (Life Technologies) was used as a donor vector. The designed AR constructs were further cloned in pDEST-HisMBP (Addgene™-Appendix). This is a bacterial expression vector that produces an N-terminal fusion protein corresponding to the inserted gene and a specific N-terminal tag. The pDEST-HisMBP vector adds an N-terminal tag containing both consecutive histidine residues and maltose binding protein (HisMBP-tag). Maltose binding protein (MBP) is a soluble protein and for this reason is frequently used to promote solubility of the expressed protein of interest fused to the tag.

To enable re-combinational cloning and efficient selection of entry or expression clones, most Gateway vectors, including the ones used in this thesis, contain two att sites flanking a cassette containing the ccdB gene for negative selection and a chloramphenicol resistance gene for counter selection. After a BP or LR recombination reaction, this cassette is replaced by the gene of interest to generate the entry clone and expression clone, respectively (Fig. 3-1). The ccdB protein is lethal for most E. coli strains. Cells that take up unreacted vectors carrying the ccdB gene or the by-product molecules retaining the ccdB gene will consequently fail to grow. This allows high-efficiency recovery of the desired clones.

To be suitable as substrate in a Gateway BP recombination reaction with a donor vector, the PCR product of the gene of interest needs to contain attB sites.

3.1.2 AR construct preparation

Two AR constructs (Fig. 3-2) were designed for biophysical studies of the AR-4Q corresponding to AR residues 1-136 and 25Q corresponding to the AR residues 1-156. These were cloned into bacterial expression vectors using the Gateway technology.

The protein constructs were designed to include an N-terminal fusion tag (HisMBP-tag), followed by a TEV protease cleavage site (amino acids ENLYFQG) to remove the tag during purification and the AR residues corresponding to the designed constructs. The TEV cleavage site was introduced in the PCR product before inserting the DNA into the donor vector.

Due to the instability of length of the polyQ in the 25Q construct and the importance of stable length of the polyQ in biophysical characterizations of the proteins, we decided to order the plasmid of 4Q and 25Q in the pDEST-HisMBP vector (GeneArt®). During the preparation of the 4Q, the gene of interests in the pDONR were transformed into OmniMAX 2 T1 phage-resistant cells (1 µL Plasmid to 20 µL cells) and plated on LB agar plates (Life Technologies). The plates were incubated for 16 hours at 37°C after which they were stored at 4°C until individual colonies were picked and grown in 5mL Lennox L Broth (LB) medium (Melford) (16 hours at 37. and 220 rpm). After 16 hours of growth, DNA was extracted from these saturated cultures by using the NucleoSpin® Plasmid purification kit (Macherey-Nagel, miniprep). The concentration of the extracted DNA was measured with a NanoDrop Spectrophotometer and the purity was estimated based on the 260/280 ratio. Afterwards the DNA was sequenced (GATC Biotech).

The pDONR Plasmid containing the gene of interest was further cloned into pDEST-HisMBP vectors via the LR recombinant reaction, following the Gateway protocol. Similarly, the reaction components (see below) were mixed, vortexed and spun down. The reaction was placed at 25°C for 4-5 hours. The reaction was terminated by adding 1 µL of Proteinase K solution, and further incubated for 10 min at 37°C. Following the incubation; the LR products were transformed into OmniMAX cells (1 µL LR reaction to 20 µL cells) and plated on LB agar plates containing 100µg/mL ampicillin (Melford). The plates were incubated for 16 hours at 37°C, and stored at 4°C until individual colonies were picked and grown in 5mL LB medium with ampicillin for selection. After 16 hours of growth, DNA was extracted from these saturated cultures using the NucleoSpin® Plasmid purification kit. The concentration of the extracted DNA was measured with a NanoDrop Spectrophotometer and the purity was estimated based on the 260/280 ratio. Afterwards, the DNA was sequenced (GATC Biotech). The sequenced expression vectors obtained from the minipreps of the LR reaction products were stored at -20°C This DNA was further transformed into the Rosetta(DE3)pLysS *E.coli* expression strain (purchased from the PECF, IRB Barcelona) for the expression of the various protein constructs. Rosetta(DE3)pLysS

cells are also commercially available from Novagen®/Merck and will be referred to as “Rosetta cells” from now on.

Cells were thawed on ice and transferred to autoclaved eppendorfs. Depending on the concentration, 1 to 2 μ l DNA (pDEST-HisMBP clone) was added to one aliquot of cells. After incubation of 30 min on ice the DNA was internalized by a heat shock of 30 seconds at 42°C, after which the cells were placed back on ice. 1mL of autoclaved LB medium was added and cells were further incubated for 1 h at 37°C and 500 rpm. After incubation, the cells were spun down gently (3 min on the bench rotor), the supernatant was removed and the cells were gently resuspended in the remaining supernatant. The resuspended cells were plated on selective LB agar plates (for OmniMAX cells: 100 μ g/mL ampicillin for pDEST-HisMBP clones; for Rosetta cells: 100 μ g/mL ampicillin + 50 μ g/mL chloramphenicol (Melford). Colonies on the plates were left to grow for 16 hours at 37°C.

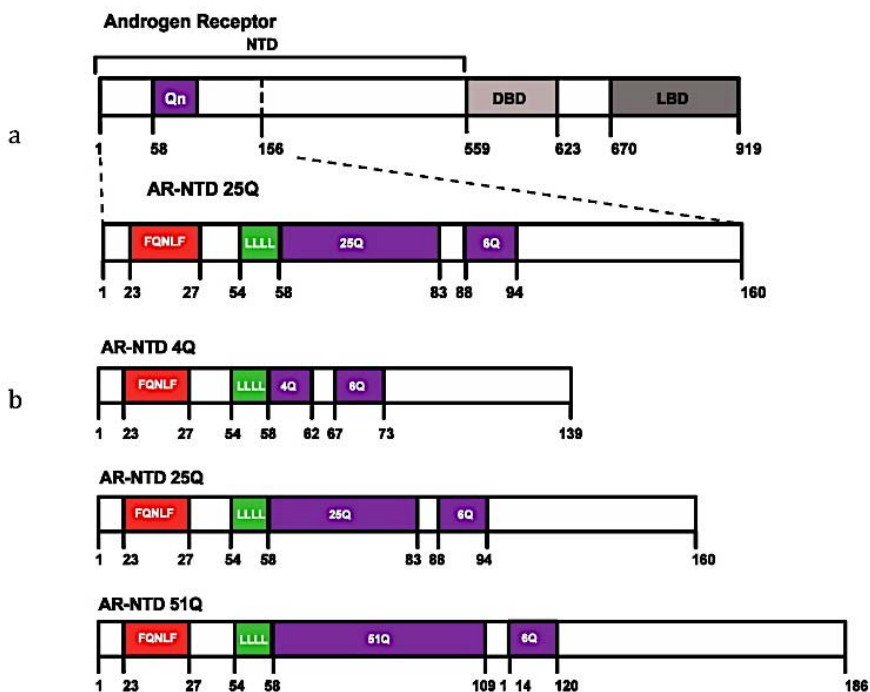


Figure 3-2: Schematic picture of full-length human AR and the constructs of interest. a) Full-length AR and respective AR-25Q construct from N-terminal domain of AR was indicated. b) AR-4Q, AR-25Q and AR-51Q constructs with the critical motifs of ²⁴FQNLF²⁸, ⁵⁵LLLL⁵⁸ and adjacent polyQ with different lengths have been indicated.

3.2. Protein expression and purification

Proteins were purified by fast protein liquid chromatography (FPLC) using an AKTA Purifier or AKTA Explorer system (GE Healthcare). Expression and purification protocols of His-tagged AR-4Q and AR-25Q starting from pDEST-HisMBP were developed and optimized in our lab (Fig. 3-3). For the production of non-isotopically labeled proteins, LB medium was used, whereas the production of uniformly single (^{15}N) or double (^{13}C , ^{15}N) isotopically labeled proteins was performed in minimal MOPS medium to which $^{15}\text{NH}_4\text{Cl}$ was added as nitrogen source and ^{13}C -glucose (double isotopically labeled sample) or unlabeled (single isotopically labeled sample) glucose as carbon source.

The polyQ fragment was cloned in a pDEST-HisMBP vector, obtaining a fusion protein containing a His-Tag, commonly used for purification and a Maltose Binding Protein, which is used to increase the solubility of the polyQ fragment. This recombinant protein was expressed in Rosetta *E.coli* cells, grown in LB at 37°C until the OD bacteria reached 0.7 and then were induced with 0.5mM Isopropyl β -D-1-thiogalactopyranoside (IPTG) for 3.5hs at 28°C.

Domain	Sequence	MW (Da)	
4Q	GMEVQLGLGR	VYPRPPSKTY	14.513
	RGAFQNLFQS	VREVIQNPGP	
	RHPEAASAAP	PGASLLLLLQQ	
	QQETSPRQQQ	QQQGEDGSPQ	
	AHRRGPTGYL	VLDEEQQPSQ	
	PQSALECHPE	RGCVPEPGAA	
	VAASKGLPQQ	LPAPP	
25Q	GMEVQLGLGR	VYPRPPSKTY	17.203
	RGAFQNLFQS	VREVIQNPGP	
	RHPEAASAAP	PGASLLLLLQQ	
	QQETSPRQQQ	QQQQQQQQQQ	
	QQQQQQQQQQ	QQQQGEDGSPQ	
	QAHRRGPTGY	LVLDEEQQPS	
	QPQSALECHP	ERGCVPEPGA	
	AVAASKGLPQ	QLPAPP	

Table3.1: Sequence of both NTD construct of AR after TEV cleavage of the HisMBP tag.

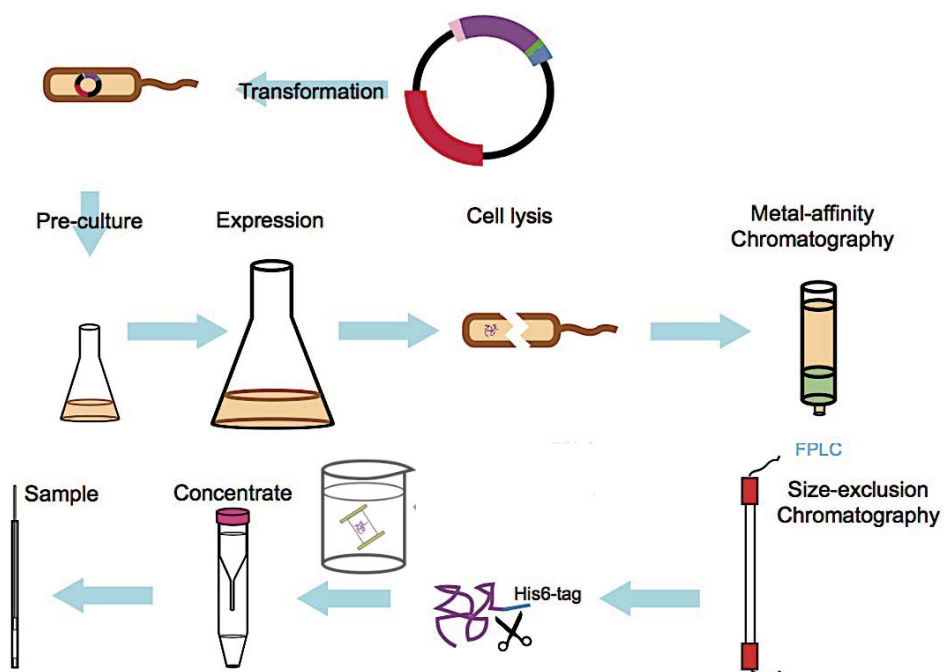


Figure 3-3: Summary of molecular Biology methods for protein expression and purification (Adapted with modification from Dr. Albert Escobedo).

3.2.1 Expression of isotopically labeled protein

The pDEST17-clone containing the AR construct of interest was transformed into Rosetta cells and plated on an LB agar plate containing 100µg/mL ampicillin for selection, 35µg/mL chloramphenicol and glucose 1%w/v. The colonies were grown for 16 hours at 37°C, after which the plate was stored at 4°C until a single colony was picked to inoculate 5mL autoclaved LB medium containing 100 µg/mL ampicillin and 35µg/mL chloramphenicol.

Alternatively, a glycerol stock of the pDEST17-clone in Rosetta cells from a previous expression was used to inoculate 5mL autoclaved LB medium containing the two antibiotics. The LB pre-culture was incubated for 16 hours at 37°C and 250 rpm. Minimal MOPS medium to which ¹⁵NH₄Cl was added as nitrogen source and ¹³C-glucose (double isotopically labeled sample) or unlabeled (single isotopically labeled sample) glucose as carbon source was inoculated 1/50 with the saturated LB pre-culture and grown for 16 hours at 37°C and 220 rpm to obtain a saturated MOPS pre-culture. A glycerol stock was prepared from the saturated LB pre-culture by adding 625µL autoclaved 80% glycerol to 375 µL cells (final concentration glycerol is 50%), and was stored at -80°C. After 16 hours of incubation of the MOPS pre-culture, minimal MOPS medium containing 100 µg/mL ampicillin and 35 µg/mL chloramphenicol, was inoculated 1/30 with the saturated MOPS pre-culture. Cells were grown at 37°C and 220 rpm to optical density (OD) measured at 600nm of 0.7 and induced with isopropyl-D-1-thiogalactopyranoside (IPTG) (final concentration 0.5mM). Induced cultures were further incubated for 4 hours at 28°C and 220 rpm. After 4 hours of incubation, the cell cultures were centrifuged (Beckman Avanti J-25 centrifuge, rotor JA-10, 6000 g, 4°C, 20 min), pellets were resuspended in lysis buffer (the recipe in following) and protease inhibitor cocktail (PIC, Sigma) was added (500 µL PIC to resuspended cells from 2 L expression). The resuspended pellets were stored at -80°C until purification.

<i>MOPS minimal medium</i>	<i>MOPS medium was prepared based on the protocol by (182)Neidhardt et al.</i>
(MOPS = 3-(N-morpholino)propanesulfonic acid, 4 morpholinepropanesulfonic acid) For the preparation of 500mL MOPS minimal media, the following components were mixed:	<ul style="list-style-type: none"> • 440mL autoclaved milliQ water • 50mL of 10X MOPS • 5mL of 0.132M K₂HPO₄ • 5mL of filter-sterilized glucose solution (7.56 g glucose in 21mL milliQ water) • 2.5mL of ¹⁵NH₄Cl filter-sterilized solution (1.05 g ¹⁵NH₄Cl in 10.5mL milliQ water) • 500 µL chloramphenicol (stock: 35 mg/mL) • 500 µL ampicillin (stock: 100 mg/mL)

Core buffer	Lysis buffer
50mM NaH ₂ SO ₄	1:1 Core buffer
500mM NaCl	1: 500 PIC
5% Glycerol	1: 500 PMSF
1mMβ-mercapto-ethanol	1:1000 Lysozyme
pH 8.0	1:2000 DNase

The 10X stock of MOPS minimal medium was prepared as follows.

1. Add 83,72 g of MOPS (Sigma, M1254-250G) and 7.17 g tricine (Sigma) to ca. 300mL autoclaved milliQ water in a 1 L beaker with a stir bar.
2. Add 10M KOH to a final pH of 7.4 and adjust the total volume to 440mL by adding autoclaved milliQ water.
3. Prepare a solution of 10mM FeSO₄·7H₂O (28 mg FeSO₄·7H₂O in 10mL autoclaved milliQ) and add it to the MOPS/tricine solution.

4. Add the following solutions to the MOPS/tricine/FeSO₄·7H₂O solution in the indicated order.

10mL of 0.276M K₂SO₄

0.25mL of 0.02M CaCl₂·2H₂O

2.1mL of 2.5M MgCl₂

100mL of 5M NaCl

0.2mL of micronutrient stock (see below)

447mL of autoclaved milliQ water

5. Filter-sterilize under sterile conditions, using a filter with pore size 0.22 μm.

6. Store in 500mL volumes at room temperature.

The 0.132M K₂HPO₄ stock solution was prepared by solubilizing 23 g K₂HPO₄ in 1L MilliQ water. The stock solution was autoclaved and stored at room temperature. The stock solution of glucose was prepared just before use by solubilizing 7.56 g glucose (¹³C-labeled or unlabeled) in 21mL autoclaved milliQ water (for 2L expression) and sterilizing it by passing it through a filter with pore diameter 0.22 μm. Likewise, the stock solution of ¹⁵NH₄Cl was prepared just before use by solubilizing 1.05 g ¹⁵NH₄Cl in 10.5mL autoclaved milliQ water (for 2 L expression) and sterilizing it by passing it through a filter with pore diameter 0.22 μm.

3.2.2 Protein purification

3.2.2.1 Ni²⁺ affinity chromatography

The frozen resuspended cell pellets were thawed and incubated with Lysozyme and DNases (Sigma, 25μL to cell lysate from expression of 2L culture) for 2 hours at 4°C turning on a wheel. After incubation with DNases, the cells were lysed by sonification (an electric sonicator for 20 min), Protease inhibitor phenylmethylsulfonyluoride (PMSF: Melford) and protease inhibitor cocktail (PIC) was added immediately to the lysed cells (500μL at 100mM PMSF to cell lysate from 2L expression). Lysed cells were centrifuged (Beckman Avanti J-25 centrifuge, rotor JA-25.50, 25000 rpm, 4°C, 30 min) to separate the soluble from the insoluble fraction. The fusion proteins were expressed as soluble proteins and were accumulated in the supernatant, which was then filtered before loading on a HisTrap HP 5mL column (GE Healthcare) for Ni²⁺ affinity chromatography. The column was equilibrated in lysis buffer and proteins samples were loaded at a rate of 3mL/min. The His-tagged fusion proteins were retained on the column. After loading the sample, lysis buffer was passed at 5mL/min (the double volume of the

loaded sample volume) to remove unspecifically bound species. Subsequently, the protein was eluted with an imidazole gradient (1mM imidazole to 500mM imidazole over 60 column volumes (CV), i.e. from 0% imidazole in elution buffer to 100% imidazole. The HisTrap HP 5ml column was used to purify the protein. This is a pre-packed column with Ni²⁺ Sepharose for preparative purification of the His-tagged recombinant proteins. As Ni²⁺ Sepharose is charged, it will selectively retain any protein, which could make a suitable complex with Ni²⁺. Therefore, the proteins with extra Histidine will have a better affinity to attach to the column and will stay in the column, while impurities will pass through the column. An additional elution step is needed to release the attached protein from the column using a solution carrying a competitor for this affinity, which in this case is imidazole. Therefore, the concentration of the imidazole is crucial during the elution step. Imidazole will attach to Ni²⁺ with a higher affinity and cause His-tagged protein to be released and come to the sample collector. The following tables have the information about the buffers and conditions, in which chromatography was performed.

Injection	
Flow rate	1ml/min
Fractionation	50ml
% Imidazole	0 (3mM)
ml of run	100 (50+50)
Wash	
Flow rate	1.5ml/min
Fractionation	50ml
% Imidazole	4% (11mM)
Elution	
Flow rate	1.5ml/min
Fractionation	2ml
% Imidazole	4→100%
mL of run	100ml

Buffer A	Buffer B
Core buffer	Core buffer
3mM Imidazole	500mM Imidazole

Column specifications	
Max flow	5ml/min
Max pressure	0.3mPa
Bed volume	5.17mL



3.2.2.2 Size exclusion chromatography

Size-exclusion chromatography (SEC) (Fig. 3-4) uses porous particles to separate molecules of different sizes. It is generally used to determine molecular weights and molecular weight distributions of polymers. Molecules that are smaller than the pore size can enter the particles and therefore have a longer path and longer transit time than larger molecules, which cannot enter the particles. Using this method additionally to the Ni²⁺ affinity, we can purify the protein based on the size. Usually, the protein samples coming from Ni²⁺ column should go through the size exclusion column to improve the purity of the protein of interest. Size exclusion chromatography with Superdex 75 and Superdex 200 was performed, using the buffers and conditions as shown below. Comparing the Ni²⁺ affinity products with size exclusion samples usually shows that there is a significant decrease in the impurities of the sample approved by Coomassie gel staining.

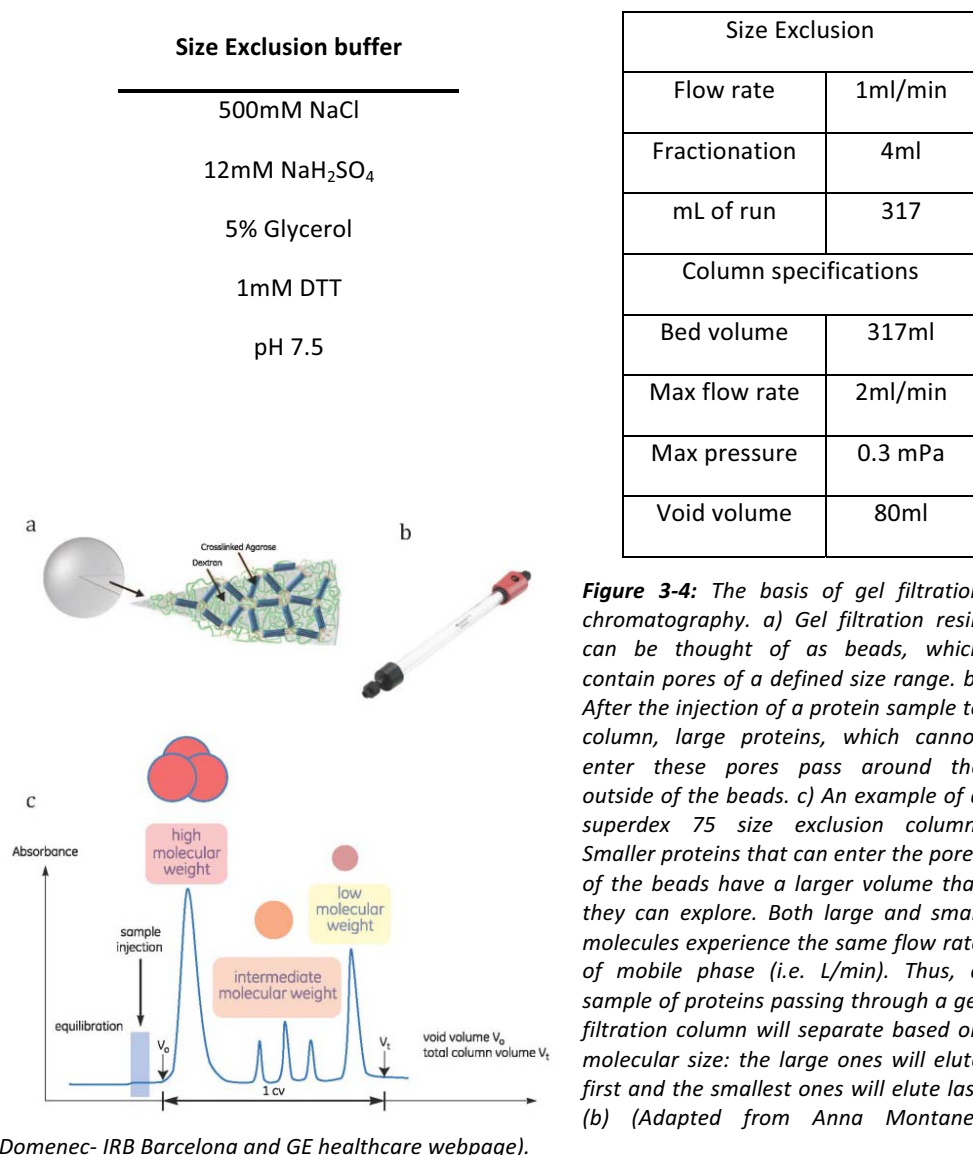


Figure 3-4: The basis of gel filtration chromatography. a) Gel filtration resin can be thought of as beads, which contain pores of a defined size range. b) After the injection of a protein sample to column, large proteins, which cannot enter these pores pass around the outside of the beads. c) An example of a superdex 75 size exclusion column. Smaller proteins that can enter the pores of the beads have a larger volume that they can explore. Both large and small molecules experience the same flow rate of mobile phase (i.e. L/min). Thus, a sample of proteins passing through a gel filtration column will separate based on molecular size: the large ones will elute first and the smallest ones will elute last (b) (Adapted from Anna Montaner

Domenech- IRB Barcelona and GE healthcare webpage).

3.2.2.3 TEV cleavage and Reverse Nickel

The pure protein was then incubated with TEV protease and dialyzed over night in an appropriate buffer (20mM NaH₂PO₄, 100mM NaCl, 0.5mM EDTA). The solution was then dissolved in 6M Urea, to dissolve the aggregates and allow a proper purification step. A further Nickel Affinity chromatography was performed. Since the protein does not contain anymore the His-Tag, it will be collected into the flow through of the column while the column will retain the used TEV protease, the cleaved HisMBP and the uncleaved protein. The product of this final step of purification was either dialyzed to sodium phosphate (20mM) for NMR experiments or dialyzed to water.

3.3. Isothermal titration calorimetry (ITC)

3.3.1 Principles and instrumentation

Isothermal titration calorimetry (ITC) is a biophysical technique that allows the simultaneous measurement of several thermodynamic parameters associated with any intermolecular interaction, including protein-ligand binding (183). The association constant (K_A , μM^{-1}), the reaction stoichiometry (n) and the binding enthalpy (ΔH , kJ mol^{-1}) are directly measured with no need of labeling or chemical derivatization of neither the protein nor the ligand (184). Using Equations 3.1 and 3.2, the values of the dissociation constant (K_D , μM , often referred to as the affinity constant), the Gibbs energy (ΔG , kJ mol^{-1}) and the entropy (ΔS , kJ K^{-1}) can be deduced.

$$3.1 \quad K_A = \frac{1}{K_D}$$

$$3.2 \quad \Delta G = \Delta H - T\Delta S$$

The instrument is composed of two equivalent cells (Fig. 3-5) made of a highly efficient thermal conducting material, typically gold. Each cell is equipped with a heater. A protecting adiabatic jacket avoids thermal exchange with the outside. The temperature of both cells is measured simultaneously. One of the cells is filled with buffer and used as a reference. Its heater applies a constant power to maintain the temperature stable along the experiment. The other cell is filled with the protein sample dissolved in the same buffer, and its heater initially applies the same power.

The experiment is performed at a constant temperature by titrating one binding partner (called the 'titrant', e.g. L) into a solution containing the other binding partner (called the 'titrand', e. g. M) in the sample cell of the calorimeter. After each addition of a small aliquot of L, the heat released or absorbed in the sample

cell is measured with respect to a reference cell filled with buffer. The heat change is expressed as the electrical power (J s^{-1}), which is required to maintain a constant small temperature difference between the sample cell and the reference cell, which are both, placed in an adiabatic jacket. Addition of L is automated and occurs from a precision syringe driven by a computer-controlled stepper motor. The contents of the sample cell are stirred to effect rapid mixing of the reactants. In commercially available instruments, volumes of sample cells are in the range 0.2–1.4ml. The amount of titrant required per experiment depends on the magnitude of the heat change; 10–100 nmol of protein are typical.

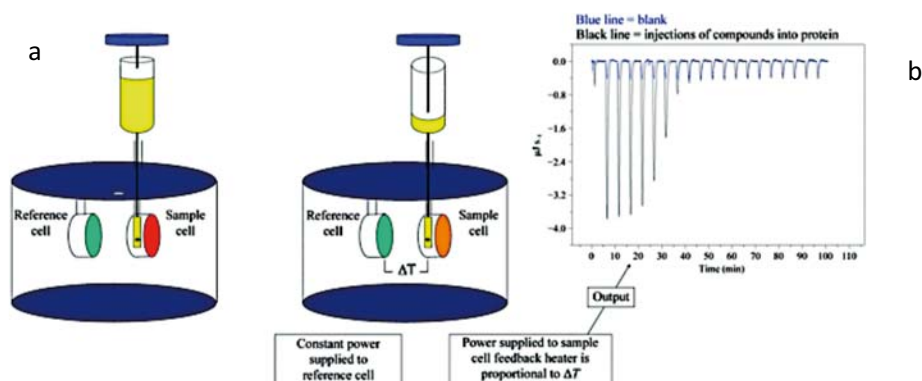


Figure 3-5: a) Schematic diagram of an isothermal titration calorimetric instrument, consisting of a sample cell that contains one binding component, an automated syringe that contains the other binding partner and a reference cell. b) Differences between the sample-cell and reference-cell temperatures induced by binding are initially translated to the power needed to bring the two samples back to temperature, before conversion to a binding enthalpy in molar terms. (Adapted from International Union of Crystallography 2007).

3.4. Fluorescence polarization substrate competition (FPSC)

Fluorescence polarization (FP) is a homogeneous method that allows rapid and quantitative analysis of diverse molecular interactions and enzyme activities. This technique has been widely utilized in clinical and biomedical settings, including the diagnosis of certain diseases and monitoring of therapeutic drug levels in body fluids. Recent developments in the field have been symbolized by the facile adoption of FP in high-throughput screening (HTS) and small molecule drug discovery of an increasing range of target classes. This method is also being widely used in studies about the protein-protein interactions, especially for proteins with known affinities to each other, like Hsps and their binding partners.

In my work, we used this system to study the binding affinity of Hsp72 and Hsp40 to AR with 2 different sizes of polyQ. The solutions of proteins with known concentrations in PBS were prepared and the binding affinities of our proteins of interest were compared with known binding affinities in this case with tau protein and a peptide from tau protein with identified binding affinity to Hsp72. Figure 3-6

(Adapted from SGC web page) is a schematic picture of fluorescence polarization assay. In this assay the peptide is typically end-labeled with fluorescein, which has a fluorescence lifetime and allow the changes in polarized signal to be detected over a wider range of molecular masses. In cases in which the peptides are known to bind to proteins, such as those in our experiments, interacting with tau protein or Hsp72 substrate binding domain, the optimum FP signal can be used to obtain information on the binding affinity of other proteins or peptides in a competition assay with the known interacting peptide.

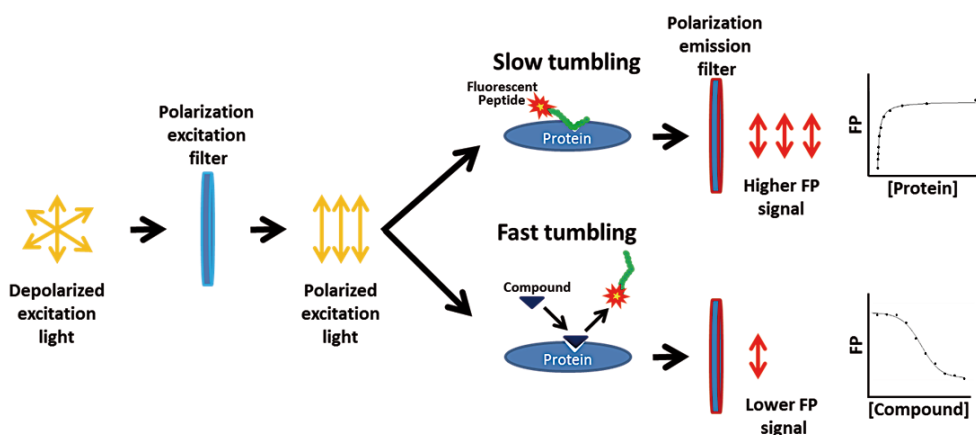


Figure 3-6: Fluorescence Polarization Assays (Adapted from SGC web page: Binding assay Using peptide displacement monitored by Fluorescence Polarization).

3.5. NMR spectroscopy of Biomolecules

3.5.1 Theoretical and practical aspects

In this section, the basic theoretical and practical concepts behind the NMR experiments that are the basis for the results presented in this thesis are summarized. An extensive theoretical description of NMR can be found elsewhere, both for initiated and non-initiated readers (185). Practical aspects of NMR are extensively described in the literature (186, 187).

3.5.2 Basic NMR characterization of a protein: 1D and HSQC

The acquisition of a ^1H 1D experiment is usually the first step in the characterization of a protein sample. This experiment does not require the isotopic enrichment of the sample and gives information on the protein fold and stability. In a folded protein, protons sample a large variety of well-defined chemical environments, so signals in the spectrum are dispersed in a wide range of frequencies ranging from 9 to almost 0 ppm. Unfolded proteins yield 1D spectra with much lower signal dispersion, which is also seen in NMR spectra of intrinsically disordered protein (IDP), as previously discussed. However, the

number of protons in a protein is large and many signals cannot be resolved. Therefore, the complexity of 1D experiments is too high to obtain further information.

3.5.2.1 [^1H , ^{15}N]-HSQC

The [^1H , ^{15}N]-HSQC (Heteronuclear Single Quantum Correlation) experiment provides a fingerprint spectrum of the protein and is the NMR experiment that was most frequently used in this thesis. Isotopic enrichment of the protein with ^{15}N allows the acquisition of the HSQC experiment. This heteronuclear NMR experiment correlates the amide proton and amide nitrogen of each amino acid in the protein backbone. It involves excitation of the proton (^1H), followed by transfer of magnetization to the ^{15}N heteronucleus and subsequent reverse polarization transfer to the proton for detection, as it is highly advantageous from a sensitivity point of view to detect magnetization on the proton (Fig. 3-7). In the resulting [^1H , ^{15}N]-HSQC spectrum, a cross peak is visible for every amino acid of the protein (construct) under study, except for proline residues and the amino-terminal residue. Proline side chains do not contain an amide proton and therefore do not have a corresponding peak in an [^1H , ^{15}N]-HSQC spectrum. The same is true for the amino-terminal residue of the protein (construct) that contains an NH^{+3} group at physiological pH. Therefore, a peak is obtained for each amino acid's backbone amide group (except for Pro), as well as for the amino groups in Lys and the aromatic rings of His and Trp; the amide groups in the side-chains of Asn and Gln; and the guanidinium group in the side-chain of Arg. Consequently, each peak is considered to roughly represent the chemical environment of each amino acid in the sequence.

Each cross peak corresponds to the chemical shifts of the amide proton and amide nitrogen from a particular residue of the protein backbone. Backbone assignment of the AR-25Q and AR-4Q constructs used in this thesis revealed the identity of the corresponding amino acid for every observed peak in the [^1H , ^{15}N]-HSQC spectra.

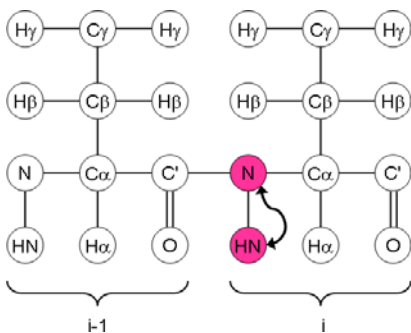


Figure 3-7: Illustration of a [^1H , ^{15}N]-HSQC experiment. The backbone of the peptide is represented. In red, magnetization is transferred from hydrogen to attached ^{15}N nuclei via the J-coupling. The chemical shift is evolved on the nitrogen and the magnetization is then transferred back to the hydrogen for detection (Adapted from (188)).

Similar to the 1D experiments, the HSQC spectrum of a folded protein displays well dispersed and sharp signals in the 6 – 10 ppm range in the ^1H dimension and

in the 135 –105 ppm range in the ^{15}N dimension, whereas this is not the case for IDP with more signal overlaps.

HSQC experiments are very useful to characterize the interaction of a protein with another one, a peptide or a small ligand. In general, a series of HSQC experiments are acquired to monitor the protein resonances while titrating progressively increasing amounts of the interacting partner. Ligand binding will cause changes in the electronic environment of residues directly involved in the interaction and those also get involved in structural rearrangements. These changes will be reflected in the spectra as chemical shift perturbations (CSP) and/or peak signal intensities changes.

The HSQC experiment has been extensively used in this thesis to characterize the protein- protein interaction in studies of Hsp40/Hsp72 with AR.

Uniformly single ^{15}N -labeled protein samples were used to measure [$^1\text{H}, ^{15}\text{N}$]-HSQC experiments. For interaction studies, the ^{15}N -labeled partner was observed by NMR, whereas the binding partner was not labeled with ^{15}N and therefore NMR-silent.

3.5.3 Sequential assignment of backbone resonances

Both the interpretation of the HSQC epitope-binding experiment and NMR time resolved experiments in identification of AR oligomerization process, requires the assignment of each resonance in the spectrum to a particular residue in the protein sequence. Due to high number of prolines in our constructs (20 Prolines), the sequential assignment of 4Q and 25Q were performed using ^{13}C direct experiments.

3.5.3.1 ^{13}C -direct NMR experiments

All ^{13}C -direct NMR experiments were done in CERM European NMR facility in collaboration with Prof. Isabella Felli. The strategy exploits the use of ^{13}C direct detection both to avoid the problem of amide proton exchange as well as to increase the chemical shift dispersion in the direct dimension of the experiments. The superior chemical shift dispersion of the heteronuclei with respect to protons is exploited by exclusively labeling the chemical shifts of the backbone's heteronuclei, using protons only as a starting source of magnetization to increase the overall sensitivity of the experiments. The resolution of the experiments has been maximized correlating five different frequencies in five different dimensions, all built up using long evolution times exploiting the favorable relaxation properties of IDPs. To substantially reduce the experimental time and maximize resolution, non-uniform sampling (NUS) was employed in all the proposed experiments. Since data sampled at not uniform intervals cannot be treated with the Fast Fourier transform (FFT), the 5D experiments were processed using the Sparse Multidimensional Fourier Transform (SMFT) algorithm (Fig. 3-9) (189). The features of this type of processing greatly simplify the visualization of the

multidimensional spectra, since the analysis of the 5D spectra is reduced to the inspection of a series of 2D cross-sections extracted from the full spectra at some fixed triads of frequencies.

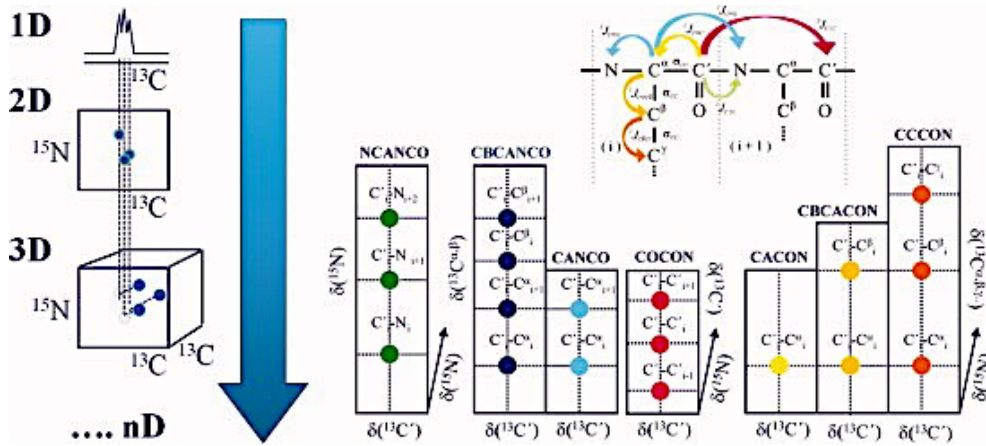


Figure 3-8: The various scalar couplings that can be exploited to design multidimensional NMR experiments, as well as the correlations expected in several ^{13}C -detected exclusively heteronuclear experiments are shown on the right panel. The increase in resolution by progressively expanding the dimensionality of NMR experiments is schematically depicted on the left. Adapted from (190).

We used the CON experiment in parallel to HSQC to obtain chemical shift information on Prolines. The CON experiment (Fig. 3-8) is a heteronuclear experiment that results in a spectrum with cross peaks for the backbone amide nitrogen ^{15}N and the carbonyl ^{13}C O of the preceding residue. Since the detection of magnetization in this experiment occurs on carbon (^{13}C), the sensitivity is low and thus longer acquisition times are required. This experiment is especially useful as cross peaks are obtained for proline residues, which are not detectable in $[\text{}^1\text{H}, \text{}^{15}\text{N}]$ -HSQC spectra as explained before. Therefore, it allows the determination of the amide nitrogen chemical shift of proline residues.

3.5.3.2 4Q construct samples preparation for assignment

A 0.4mM sample of $^{13}\text{C}, ^{15}\text{N}$ double-labeled 4Q construct in 20mM Sodium phosphate, 1mM TCEP, pH 7.4 was prepared. 10 % D_2O was added for the lock. The sample was divided in several aliquots that were used to acquire the NMR experiments for the resonance assignment and those to follow the temperature dependence. An identical sample, but exclusively enriched in ^{15}N , was used to acquire the ^{15}N relaxation experiments.

3.5.3.3 Resonance assignment

In collaboration with Alessandro Pia in CERM European NMR facility in Florence, all the multidimensional NMR experiments tailored for sequence-specific assignment were acquired at 278 K. ^{13}C -detected 4D HCBCACON, 4D HCBCANCO,

4D (HCA)CON(CA)CON and 4D (HN)CON(CA)CON experiments were acquired at 16.4 T on a Bruker Avance spectrometer operating at 700.06 MHz ^1H , 176.03 MHz ^{13}C and 70.94 MHz ^{15}N frequencies, equipped with a cryogenically cooled probehead optimized for ^{13}C -direct detection (Fig. 3-8). ^1H - detected 4D TROSY HNCO, 4D TROSY (H)NCO(CA)NNH and 4D TROSY HN(COCA)NNH experiments were performed at 22.3 T on a Bruker Avance III spectrometer operating at 950.20 MHz ^1H , 238.93 MHz ^{13}C and 96.28 MHz ^{15}N frequencies, equipped with a cryogenically cooled probehead. On-grid non-uniform sampling (NUS) was employed to increase the spectral resolution and to speed up the acquisition of the experiments. The parameters used for the acquisition of the experiments are reported in the Appendix.

3.5.3.4 Temperature dependent experiments

A series of 2D ^1H - ^{15}N HSQC and 2D CON-IPAP spectra were collected in the range 278-308 K, with steps of five degrees, at 16.4 T on a Bruker Avance spectrometer operating at 700.06 MHz ^1H , 176.03 MHz ^{13}C and 70.94 MHz ^{15}N frequencies, equipped with a cryogenically cooled probe head optimized for ^{13}C -direct detection. The parameters used for the acquisition of the experiments are reported in the Appendix.

3.5.3.5 Backbone dynamics experiments

^{15}N relaxation experiments (R_1 , R_2 and $^{15}\text{N}\{^1\text{H}\}$ NOEs) were performed at 278 K at 16.4 T on a Bruker Avance spectrometer operating at 700.13 MHz ^1H , 176.05 MHz ^{13}C and 70.94 MHz ^{15}N frequencies, equipped with a cryogenically cooled probe head. The parameters used for the acquisition of the experiments are reported in the Appendix.

3.5.3.6 25Q construct samples preparation for assignment

A 0.4mM sample of ^{13}C , ^{15}N double-labeled 4Q construct in 20mM Sodium phosphate, 1mM TCEP, pH 7.4 was prepared. 10 % D_2O was added for the lock. The sample was divided in several aliquots, which were used to acquire the NMR experiments for the resonance assignment. Three additional ^{15}N -labeled samples of different concentrations (0.5, 0.25 and 0.125 mM) were used to acquire the ^{15}N relaxation experiments.

3.5.3.7 Resonance assignment

All the multidimensional NMR experiments tailored for sequence-specific assignment were acquired at 278 K. ^{13}C -detected 4D HCBCACON, 4D (HN)CON(CA)CON and ^1H - detected 3D HNCO experiments were acquired at 16.4 T on a Bruker Avance spectrometer operating at 700.06 MHz ^1H , 176.03 MHz ^{13}C and 70.94 MHz ^{15}N frequencies, equipped with a cryogenically cooled probeheads optimized for ^{13}C -direct detection. ^1H - detected 4D TROSY HN(CA)CO and 4D TROSY HN(COCA)NNH experiments were performed at 22.3 T on a Bruker Avance

III spectrometer operating at 950.20 MHz ^1H , 238.93 MHz ^{13}C and 96.28 MHz ^{15}N frequencies, equipped with a cryogenically cooled probeheads. On-grid non-uniform sampling (NUS) was employed to increase the spectral resolution and to speed up the acquisition of the experiments. The parameters used for the acquisition of the experiments are reported in the Appendix.

To obtain information on the backbone dynamics, ^{15}N relaxation experiments (R_1 , R_2 and $^{15}\text{N}\{^1\text{H}\}$ NOEs) were performed at 278 K at 16.4 T on a Bruker Avance spectrometer operating at 700.13 MHz ^1H , 176.05 MHz ^{13}C and 70.94 MHz ^{15}N frequencies, equipped with a cryogenically cooled probe head. The parameters used for the acquisition of the experiments are reported in the Appendix.

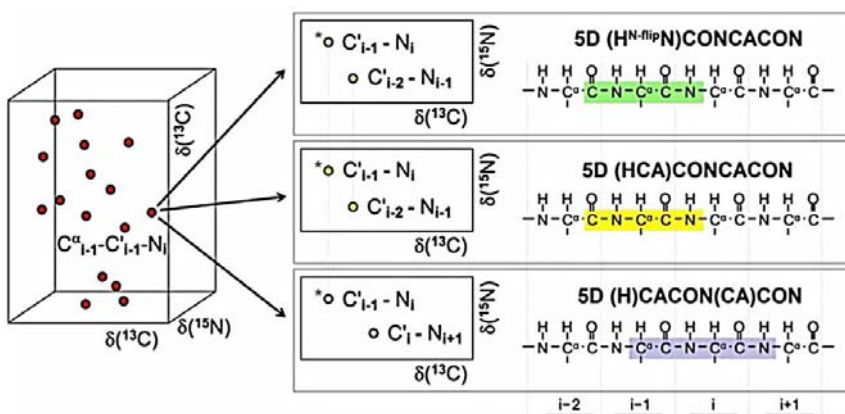


Figure 3-9: Using 5D approach for backbone assignments of IDPs. The CON peak that correlates the same $\text{C}^{i-1}-\text{N}_i$ frequencies as the associated peak in the basis spectrum is marked by an asterisk. On the right part of the figure, the nuclei involved in the coherence transfer pathways are highlighted on the backbone of the protein. Adapted from (133).

3.5.3.8 Sample preparation for interaction studies with Hsp40 and Hsp72

For NMR experiments, purified protein and peptide solutions containing 10% D₂O were adjusted to pH=7.4 when required and loaded into conventional or Shigemitsu tubes (Shigemitsu, Tokyo, Japan). 25Q and 4Q samples were prepared as mentioned in the assignment experiments. Hsp40 and Hsp72 were expressed and purified by Prof. Gestwicki's laboratory in UCSF and sent to us. The pH and buffers condition of samples were adjusted by an overnight dialysis to a compatible buffer for NMR experiments.

3.5.3.9 NMR acquisition for interaction studies between AR and Hsps

The [^1H , ^{15}N]-HSQC spectra were acquired in the Bruker 800 spectrometer with the following parameters: 2048 complex points in the F2 dimension (^1H) and 512 increments in the F1 dimension (^{15}N). The SW was set to 8013 Hz (10 ppm) (^1H) and 1784 Hz (22 ppm) (^{15}N), and the number of scans to 16. The x-carrier

frequency was determined by referencing to internal DSS-d6. The DSS-d6 frequency was obtained from a 1D experiment, recorded immediately before the 2D [¹H, ¹⁵N]-HSQC. Indirect referencing was used in the ¹⁵N dimension by the use of the conversion factors from Wishart et al (191).

3.5.4 Secondary structure propensity

Following the assignment of proteins, secondary structure propensities were characterized based mainly on their ¹³C α chemical shifts. These chemical shifts are sensitive indicators for secondary structure elements in folded proteins and residual secondary structure in IDPs.

The main method for determining secondary structure propensity of a protein is the analysis of its secondary chemical shifts. These are given by subtracting so-called "random coil" chemical shifts from the experimental chemical shifts. Random coil chemical shifts are empirically determined chemical shifts for a residue in a completely disordered conformation. Typically, those have been determined either by studying a series of short disordered peptides (137, 192) or by using available data for coil regions of proteins in the Protein Databank (PDB), the NMR chemical shift assignments of denatured proteins or IDPs that have been deposited in the BioMagResBank (BMRB) database. Since chemical shifts are strongly dependent on the type of amino acid, the random coil value of the corresponding amino acid type is subtracted from the observed chemical shift.

The C α and C β chemical shifts are strongly dependent on the amino acid type. The most common values for each amino acid type are extensively characterized. Therefore, a particular ¹H-¹⁵N resonance in the HSQC spectrum can be assigned to a particular residue in the protein sequence. The chemical shifts are strongly correlated with local structure. Thus, the values of the C α and C β chemical shifts assigned to a particular residue are informative on the secondary structure element where it is located. This information can be extracted by calculating the secondary chemical shift $\Delta\delta$ (Equation 3.3).

$$3.3 \quad \Delta\delta = \delta_{observed} - \delta_{random\ coil}$$

Thus, the observed chemical shift of a particular residue in the protein sequence is compared to the chemical shift of this residue type in a random coil. This is a reference value corresponding to the average of chemical shifts that a certain amino acid type can have in unstructured or disordered parts of a protein (random coils). Depending on the value of $\Delta\delta$, the main secondary structure elements can be discerned (193) :

- For residues located in α helices $\Delta\delta C^\alpha$ values tend to be positive, whereas $\Delta\delta C^\beta$ are generally negative. In this case, $\Delta\delta C^\alpha - \Delta\delta C^\beta$ will be positive.
- For residues located in β sheets $\Delta\delta C^\alpha$ values are most likely negative, while $\Delta\delta C^\beta$ tend to be positive. In this situation, $\Delta\delta C^\alpha - \Delta\delta C^\beta$ will be negative.

3.5.5 Relaxation and dynamics

Relaxation is the process by which coherent (processing in phase) transversal magnetization is converted into de-phased magnetization aligned with the z-axis (longitudinal). There are two relaxation mechanisms:

- Longitudinal relaxation T_1 is the name given to the re-alignment of transversal magnetization with the z-axis (B_0). Typical values found in proteins are in the range of several hundred milliseconds to a few seconds.
- Transverse relaxation T_2 consists in the de-phasing of the transverse magnetization in the x-y plane causing the loss of coherence. It is caused by stochastic fluctuations of the magnetic field B_0 that have different effects on the molecules of the sample. In proteins, it has values in the order of several tens of milliseconds.

Both T_1 and T_2 depend on molecular tumbling, described by the correlation time τ_c . To probe the dynamics of AR-25Q and AR-4Q, we recorded ^{15}N transverse relaxation (T_2) experiments for these constructs. The corresponding relaxation rates (R_2) report on time scales in which conformational transitions occur (μs to ms).

Generally, ^{15}N longitudinal and transverse relaxation rates (R_1 and R_2 , respectively) and heteronuclear NOE (hetNOE) of the backbone amides are measured to characterize the dynamical properties of a protein backbone. The ^{15}N R_2 relaxation rates were, however, found to be particularly sensitive to any deviation from the random coil (194) as they showed larger variations than the other two relaxation parameters. The very same observation on sensitivity of R_2 was seen in 25Q and 4Q relaxation experiments.

Furthermore, for several unfolded proteins, the observed variations in R_2 relaxation rates clustered in certain regions of the sequence while they seemed to be more randomly distributed in R_1 relaxation rates and hetNOE (194). Therefore, the R_2 relaxation rates are particularly informative when characterizing the motional properties of a disordered polypeptide chain.

Flexible parts of a protein relax slower than more rigid parts of the sequence and are therefore characterized by lower R_2 relaxation rates. By contrast, high R_2 relaxation rates are indicative of secondary structural elements, tertiary contacts, local collapse or other phenomena that give the polypeptide chain a certain degree of rigidity. Chemical exchange can also contribute to transverse relaxation.

Internal flexibility measurements are often complemented with the measurement of the ^1H - ^{15}N heteronuclear NOE experiment, which provides information on the motion of backbone amide N-H bonds. The relative NOE intensity between these two nuclei is obtained from two HSQC spectra with and without proton presaturation, which are acquired in an interleaved manner. Residues in folded

regions display $I_{\text{sat}}/I_{\text{insat}}$ ratio values around 0.8, while those located in flexible areas display smaller values.

3.5.6 ^{15}N Relaxation and heteronuclear NOEs

When the system investigated is simplified and the local structure does not have a significant impact on the NMR observable, as in the case of the ^{15}N auto and crossrelaxation rates, the analysis of the experimental data can highlight regions with different characteristics. For instance, the nitrogen relaxation rates ($^{15}\text{NR}_1$, $^{15}\text{NR}_2$, $^1\text{H}-^{15}\text{N}$ NOEs) can be used to identify parts of the polypeptidic chain characterized by different motional properties (195, 196). Among them, in the case of flexible systems, the heteronuclear NOE is the most sensitive reporter of fast local motions as it is primarily determined by high-frequency spectral density functions. Indeed, the heteronuclear NOEs, depending on the local effective correlation time, span a large range from -4 to $+1$ and for this reason provide very effective information on the regions of the protein characterized by different mobility. This can be combined with the analysis of transverse relaxation rates that, in addition to fluctuations depending on the local correlation time, are also sensitive to exchange processes.

3.5.7 Proton–proton NOEs

Proton–proton (H-H) NOEs are the observables with the richest structural information for the structural determination of globular proteins. Indeed, for well-folded proteins, conformational averaging and differential local dynamics affect the H-H NOEs to a minor extent, as they largely depend on an overall correlation time and can thus provide information on internuclear distances. In the case of the structural and dynamic heterogeneity typical of IDPs, this is obviously not a good approximation and the difficulty consists of separating the structural and dynamic information. In general, proton–proton NOEs are reduced in intensity due to averaging as well as to generally smaller effective local correlation times – two aspects that reduce the cross peak intensities. Moreover, they involve protons, which generally overlap extensively in IDPs and are thus not easy to assign in an unambiguous way. Various experimental solutions to this problem have been proposed in which heteronuclei are frequency labeled to reduce the ambiguities in the assignment. In general, the most intense cross-peaks in the spectra arise from intra-residue or sequential NOEs, while long range NOEs are not easy to detect but, if detected, they are very informative of partial local structured conformations present at equilibrium.

3.5.8 Chemical shift perturbation experiments

Protein-protein HSQC binding studies were performed using a constant concentration of the ^{15}N -labelled AR construct (33 μM) and adding different equivalent of Hsp72 and Hsp40. The signal intensity of the peaks was analyzed and compared in presence and in absence of binding partners. The signal intensity

loss is a direct indication of involvement of one specific motif in binding and interaction with another protein. Different equivalent of Hsp72 and Hsp40 showed a concentration dependent signal intensity loss in HSQC spectra.

The analysis of chemical shift perturbations in all experiments was performed using the CCPNMR Analysis software (197) using a 0.2 weighting of the ^{15}N dimension relative to ^1H (Equation 3.4).

$$3.4. \quad \Delta\delta = \sqrt{\frac{(\Delta\delta^1\text{H})^2 + \left(\frac{\Delta\delta^{15}\text{N}}{5}\right)^2}{2}}$$

3.6. Transgenic mouse colony maintenance

The transgenic mice Carrying Full-Length Human AR with 97 CAGs Showed Sexual Differences in Symptoms, Transgene Expression, Pathology and Symptomatic differences with gender. We transferred transgenic mice from Japan (Riken institute) expressing the full-length human AR containing 24 or 97 CAGs under the control of a cytomegalovirus enhancer and a chicken β -actin promoter (Fig. 3-10). Three lines of animals with 24 glutamines (AR-24Q) and 5 lines with 97 glutamines (AR-97Q) were established in Japan (Nagoya Medical school). Copy numbers of the transgene were 1 to 5 in AR-24Q mice and 1 to 3 in AR-97Q mice. They assessed the 24 or 97 CAG repeat in the transgene by the genotyping on the tail DNA using a fluorescently labeled primer, and a subsequent size determination on the polyacrylamide gel did not show unequivocal intergenerational instability in the CAG repeat number (198).

Three of five lines with AR-97Q (#2–6, #4–6, #7–8) exhibited progressive motor impairment, although no lines with AR-24Q showed any manifested phenotypes up to the latest age examined which was 30 weeks old. All three symptomatic lines showed a small body size, a short life span, a progressive muscle atrophy and weakness, as well as reduced cage activity, all of which were markedly pronounced and accelerated in the male AR-97Q mice, but not observed or far less severely in the female AR-97Q mice regardless of the line. The first detectable phenotype is a muscle atrophy of the trunk and hind limbs followed by weight loss and impairment of the rotarod task. The onset of motor impairment were detected by the rotarod task at 8 to 9 weeks of age in the male AR-97Q mice, while after 15 weeks or more in the female mice. The affected mice were hypoactive and dragged their hind limbs. The forelimbs were not involved until hind limb atrophy became severe. Males showed a markedly faster and earlier motor deficit than females, and shorter life span. The 50% mortality ranged from 66 to 132 days of age with the male AR-97Q mice of 3 lines, whereas mortality of the female AR-97Q mice remained only 10% to 30% at more than 210 days in the three lines. AR-24Q mice showed no pathologic abnormalities at the age of 12

weeks. In AR-97Q mice, diffuse nuclear staining and less frequent NIs were detected with 1C2, an antibody specifically recognizing the expanded polyQ over 38 (199).

3.6.1 Re-characterization of SBMA transgenic mice

In collaboration with Prof. Sobue laboratory in Nagoya Medical School, transgenic mouse models of SBMA harboring human AR with 2 different size of polyQ, called 24Q and 97Q were generated from a cryo-preservation. The animals express the full length human AR contain 24 or 97 CAGs under the control of a cytomegalovirus enhancer and a chicken beta-actin promoter (92). The AR-97Q males exhibit progressive motor impairment, while no AR-24Q mouse showed any phenotype during their life span. By direct DNA sequencing, the presence of 24 and 97 CAG repeated sequences was confirmed. All animals were genotyped by Transnetyx genotyping company® (USA). Ethical approval for these studies was obtained from ethical committee on animal experiments at the Institute for Research in Biomedicine (IRB).

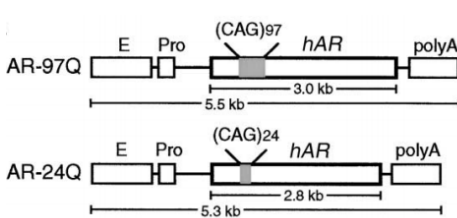


Figure 3-10: Scheme of 24Q and 97Q transgene. To generate the transgenic mouse model of SBMA, human AR with 2 different size of the polyQ, either 97 or 24 have been over expressed, using a chicken β -actin promoter. Adapted from (92).

3.6.2. Transgenic mouse embryo transfer from Japan

The transgenic mouse model of SBMA we used for our experiments has been generated by laboratory of Prof. Gen Sobue. Among different introduced transgenic lines in SBMA field, we found this model the most appropriate one for our research and the questions we wanted to answer, as thoroughly explained in the previous part (re-characterization of transgenic animals).

We received the transgenic animal embryos of 24Q (B6D2-Tg(CAG-AR*24Q)5-5Sobu) and 97Q (B6D2-Tg(CAG-AR*97Q)7-8Sobu) animals from Riken Institute in Japan. They provided us with 100 embryos from each line and also the recovery protocol. Thanks to Embryo Transfer facility in IRB, they provided us with a number of positive and negative embryos. In the following, there is a summary of the recovery and embryo transfer procedure (Fig. 3-11).

1) The temperature of the thawing media and the duration of embryo exposure to them are critically important for the survival of embryos. We compared the morphology of our embryos with that of embryos in the photos shown above. One may extend the period for the first thawing medium (TS1) until the embryos recover from a shrunken morphology.

2) The embryos just after thawing are very sensitive to physical damages. It is strongly suggested to handle them as gently as possible.

3) We started to first thaw the C57BL/6 Wt and 24Q embryos and made sure that the embryos finally look normal in the M16 medium, as the 97Q transgenes were more complicated.

4) Embryos should be transferred into day 0.5 (on the day of vaginal plug) oviducts on the day of thawing. Ideally, 16-20 embryos are to be transferred per female so that they may get pregnant efficiently. And therefore, we started with 5 females.

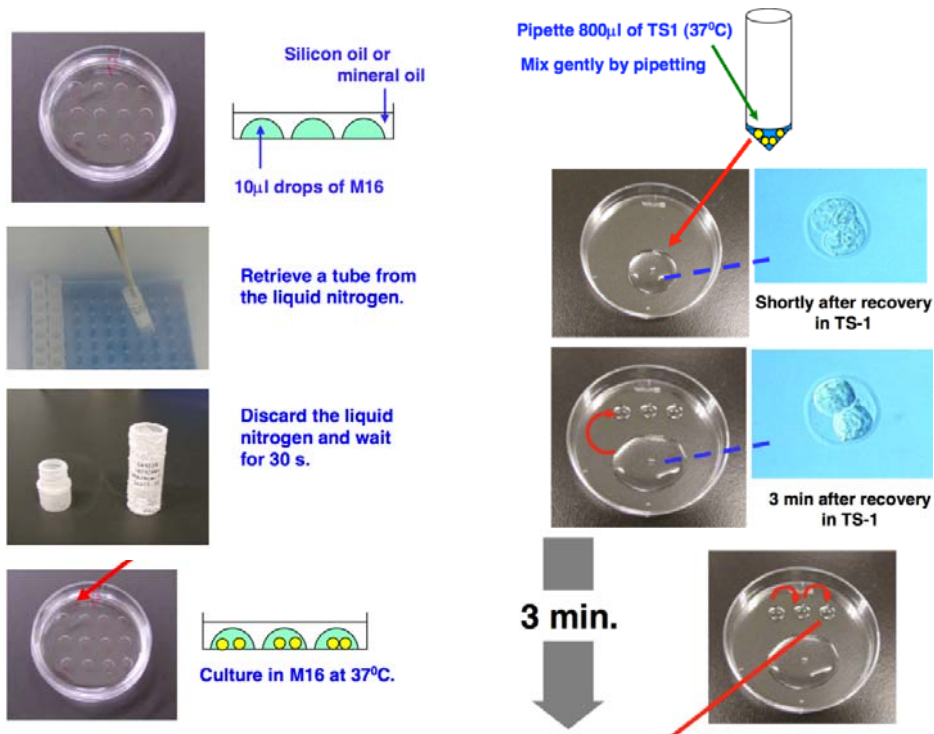


Figure 3-11: Schematic picture of the embryo transfer procedure. 1-(left) Prepare the culture dish with 10ml aliquots of M16 and warm the TS1 to 37oC. 2- Put on the facemask and cryogloves and retrieve a cryotube from the liquid nitrogen container using forceps. 3- Quickly discard the liquid nitrogen in the tube, and let it stand for 30 s at room temperature. 4- Put 800 ml of TS1 (37oC) into the tube by a single out-pipetting, and then mix the solution gently by pipetting ten times in 25 sec. Make sure that the solution is evenly dissolved. 5-Using a 1000 ml pipette, transfer the entire volume to a watch glass or 50 mm petri dish. Hereafter, embryos are handled at room temperature until they are placed in M16. 6- Set the timer for 3 min and start it. Spread the solution over the surface of the watch glass or petri dish by gentle shaking.

Dispersion of the solution makes it easier to spot the embryos. 7- After 3 min of warming, retrieve the embryos using a sterile mouth-pipette, and transfer them from the drop of TS1 to a drop of TS2. Set the timer for 3 min and start it. 8- Transfer the embryos to the second drop of TS2. Then, transfer them to the third drop of TS2. 9-After 3 minutes transfer the surviving embryos to 10 ml drops of

M16. Incubated embryos under 5% CO₂ in air at 37°C until transfer to the oviducts.

3.6.3 Transgenic mouse breeding plan

To start with the breeding strategies, we had to take 4 points into account (Fig. 3-12).

1. Age: The 97Q animals start to have phenotype by age of 9 to 10 weeks. And due to their disease progression and due to presence of the transgene, they need to be sacrificed by maximum 5 months of age (usually the average is by 3 months).

2. Number of mice: While 24Q animals follow the Mendelian female/male ratio, in 97Q animals less than 50% of the pups are male, which is the main restriction for maintaining a large enough number of positive animals.

3. Gender: Our experimental mice are only male groups, as the disease progression is Testosterone (DHT) dependent, and only male mice show the phenotype of the disease. Therefore, these selective experimental conditions clearly constrain our experimental plan.

4. Healthy male mice are a necessity for mating and breeding: As we already mentioned, the SBMA transgenic animals are very similar to patients of the disease and start to lose their male sexual activity as the disease progresses. In these mice, additionally, skeletal muscle atrophy in legs is the main reason for their inability of mating behaviors. Therefore, we considered all these points and were able to increase the number of positive 97Q and 24Q animals.

Our final breeding strategy is shown in the following scheme.

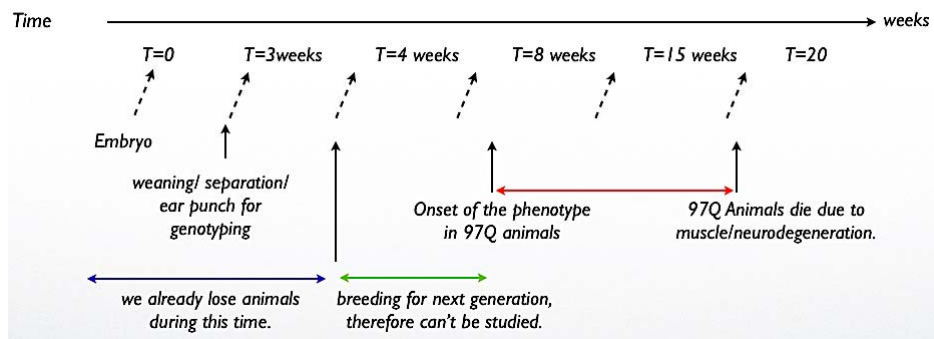


Figure 3-12. Scheme of Breeding strategy and challenges of maintaining the SBMA transgenic mouse model.

3.7. Transgenic mouse tissue homogenate characterization

3.7.1 Biochemical characterization

3.7.1.1 Tissue dissection and homogenate preparation

For biochemical characterization of transgenic animal tissue, 3 tissues of interest were dissected. According to the literature, the mainly involved tissues in SBMA pathology in this transgenic mouse model are Spinal cord and hind limb muscles. Full Brain attached to spinal cord was removed with surgical scissors and spinal cord, brain stem and frontal cortex have been separated (as shown in figure 3-13) and instantly froze on dry ice. Hind limb muscles were dissected as shown below from both legs and froze at -80°C . The samples were kept in -80°C until the day of homogenate preparation. As it is shown in figure 3-13 we used a tissue douncer and not an electric tissue homogenizer. We have found out that it is critical for the experimental condition and the idea was adapted from Prof. Sobue laboratory.

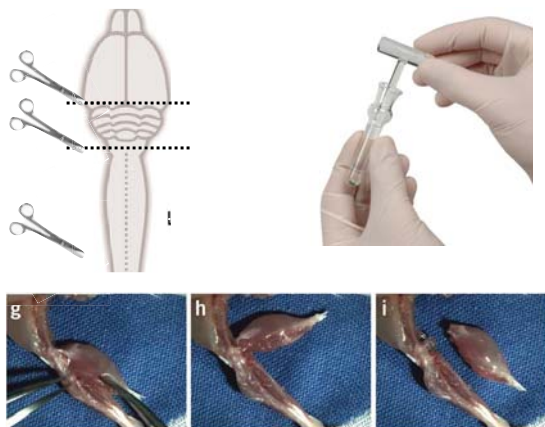


Figure 3-13. a) For biochemical and biophysical characterization of tissues involved in SBMA, spinal cord and Brain of transgenic mice was dissected and separated in 3 parts of interest. b) Hind limb muscle was separated carefully and analyzed with biochemical analysis. c) A dounce tissue grinder has been used to prepare the tissue homogenate (Adapted from (200)).

3.7.1.2 Protein expression analysis in tissue homogenate (Western blot)

We exsanguinated the mice under ketamine-xylazine anesthesia and have frozen the tissues on dry ice. Afterwards, the tissue (0.1 g wet weight) was homogenized in a cellLytic[®]M lysis reagent and homogenizing buffer (Sigma), 1mM 2-mercaptoethanol with 1mM PMSF and aprotinin at $6\mu\text{g}/\text{ml}$ and was centrifuged in 2500 g for 15 min at 4°C . Protein concentration of the prepared tissues was measured using DC[®] protein assay from Bio-Rad. Each lane on a 5%–20% SDS-PAGE gel was loaded with protein 200 μg for nervous tissue, and 80 μg for muscular tissue from the supernatant fraction, which was transferred to Hybond-

P[®] membranes (Amersham Pharmacia Biotech, Buckinghamshire, England), using 25mM Tris, 192mM glycine, 0.1% SDS, and 10% methanol as transfer buffer. After immuno-probing with rabbit anti-AR antibody H280 (1:1000) (Santa Cruz Biotechnology, Santa Cruz, California), we performed second antibody probing and detection using the ECL+plus kit (Amersham Pharmacia Biotech).

3.7.2 RNA quantification analysis

Mice were exsanguinated under ketamine-xylazine anesthesia, and we snap-froze their tissues with liquid N₂ and extracted the total RNA from the tissues with Trizol[®] (Life Technologies/Gibco[®] BRL, Gaithersburg, Maryland), and reverse transcribed the RNA using SUPERScript II reverse transcriptase (Life Technologies/Gibco[®] BRL). We used 5'-TTCCACACCCAGTGAAGC-3' and 5'-GGCATTGGCCACACCAAGCC-3' as primers for specific transgene RNA detection. qPCR was used to amplify the cDNA and the intensity of the qPCR products signals were compared with those of GAPDH mRNA levels, which have been separately amplified as a host gene control. For each experiment 4-independent qPCR have been performed.

3.7.3 Antibodies

For all biochemical studies, immunohistochemistry and immunofluorescence staining the following antibodies have been used.

H280 (Santa Cruz Biotechnology, California) is a polyclonal rabbit antibody that was raised against amino acids 91 to 370 of AR of human origin. C19 and N20 (Santa Cruz) are also both polyclonal rabbit antibodies that were raised against the very C-terminal and N-terminal regions of human ARs. F39.4.1. mouse monoclonal antibody from Biogenex[®] (Netherlands) specific for N-terminal domain of AR was used for WB analysis and immunofluorescence staining. For immunohistochemical analysis, IC2 antibody (Millipore[®]), a mouse monoclonal antibody specific for polyQ length over 38, was used. This antibody specifically detects the polyQ aggregates with higher polyQ tract numbers. For immunofluorescence staining of AR, Alexa-Flour-488 and Alexa-Flour-568 from Life technologies were used. For immunofluorescence staining, the membrane staining was done by Alexa[®] Fluor 594 *wheat germ agglutinin* (WGA) from Life Technologies.

DAPI (4',6-diamidino-2-phenylindole) is a fluorescent stain that binds strongly to A-T rich regions in DNA. It is used extensively in fluorescence microscopy for nuclear staining. As DAPI can pass through an intact cell membrane, it can be used to stain both live and fixed cells.

3.8. Quantification of aggregate amount in tissue homogenate

3.8.1 Seprion ligand for the study of Protein Aggregation Diseases

Seprion kits (Microsense Biotechnologies, UK) easily and efficiently enable the detection of abnormal aggregated protein (Fig. 3-14). Applications include the study and monitoring of protein aggregation diseases in human and animal samples and drug screening *in vitro* and *in vivo*. To date, the Seprion ligand products have been demonstrated to work with β -amyloid, tau, α -synuclein, Htt and prion proteins. Evaluation kits are available for other applications.

The Seprion ligand is available in two forms, coated onto microplates (Seprion ligand ELISA) or coated onto magnetic beads (PAD-beads®).

The Seprion-coated microplates are available in two options, SEP1 and SEP2 and are used in an ELISA format. Upon incubation of samples in the microwells, aggregated proteins (but not the normal unaggregated proteins) bind to the Seprion ligand. After washing, the immobilized aggregated protein is then detected with an appropriate antibody. For these experiments, H280 antibody specific for AR N-terminal domain was used.

The Seprion-coated magnetic beads (PAD-beads®) have extracted aggregated proteins from a larger volume of the sample. After the capture onto the bead, the aggregates have been eluted and detected in an ELISA that is specific for the protein of interest.

The Elution step can be performed afterwards by KCl, SDS 0.75% or HCl, when it is immediately neutralized by NaOH.

3.8.2 Seprion ligand ELISA

Brainstem attached to spinal cord and hind limb muscle tissues from SBMA transgenic mice and Wt mice were dissected, frozen on dry ice and stored at -80°C until required. A 10%v/w lysate was prepared in homogenizing buffer (Sigma), 1mM b-mercaptoethanol; 1mM PMSF; 1mM DTT; protease inhibitor cocktail (Roche)). The total tissue homogenate concentration was measured using DC protein assay (Bio-rad). Identical concentrations of tissues were transferred to the well of a Seprion ligand-coated ELISA plate and were incubated with shaking for 1 h at room temperature (RT). After the removal of the lysate, the well was washed 5x in PBS-T (PBS; 0.1% Tween) and 100 μ l primary antibody (diluted 1:2000 in conjugate buffer (150mM NaCl; 4% BSA (98% electrophoretic grade); 1% non-fat dried milk; 0.1% Tween 20 in PBS) was added and incubated with shaking for 1 h at RT. After five washings with PBS-T, 100 ml horseradish peroxidase (HRP)-conjugated rabbit anti-goat secondary antibody (DAKO®) (1:2000 in conjugation buffer) was added and incubated while shaking for 45 min at RT. After washing

five times with PBS-T, 100 ml of TMB substrate (SerTec) at RT was added and incubated in the dark (wrapped in foil) at RT for 30 min. Reactions were terminated by the addition of 100 ml 0.5 M H₂SO₄ and the absorption at 450 nm was measured using a plate reader (Bio-rad).

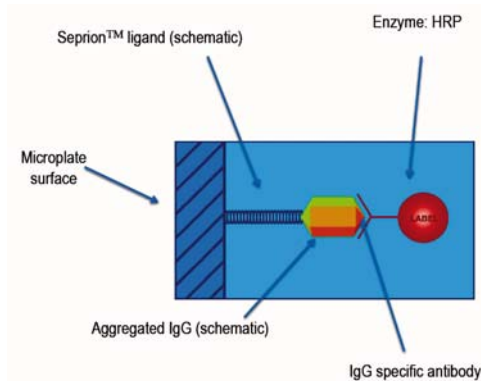


Figure 3-14: Schematic image of a Seprion ligand. High molecular weight polymeric ligand (blue) consists of repeating charged and hydrophobic chemical groups that interact with similar repeating groups that occur on aggregated proteins. The selectivity for aggregated proteins is based on the arrangement of large numbers of polar and hydrophobic regions in aggregated proteins (yellow) that occur to a much lesser extent in a single unaggregated protein molecule (Adapted from CalBioReagents® web page).

3.9. Isolation of aggregated species from ex vivo material

3.9. 1 Seprion ligand bead capture (PAD-beads)

10% w/v tissue homogenates were prepared as explained before. Seprion-coated magnetic beads of 100 µl (Microsens Biotechnologies®) were transferred to a low-bind eppendorf tube on a magnetic particle concentrator (Millipore). The supernatant was removed and replaced with 100 µl of ultrapure water prior to the use in the assay.

About 100 µl of homogenate was diluted to 700 µl with water and made up to 1 ml with 100 µl capture buffer (Microsens Technologies®) and 100 µl of coated beads. The tube was shaken on Vibrax shaker (VXR basic IKA Vibrax®) for 1 h at 700 mot/min at RT, transferred to the magnetic concentrator and the supernatant was removed. The beads were washed with 500 µl of capture buffer (wash buffer 1) followed by 1 ml of TBS (wash buffer 2)(Microsens Biotechnologies®) and 300 ml of TBS. Traces of TBS were removed by pipette after quickly spinning in a microfuge. The captured material was eluted by mixing with 20 ml of 1% SDS in case of WB analysis or with KCl and heating at 100°C for 5 min in a heating block. For a better efficiency of capturing large aggregated species by the Seprion ligands on the beads, the shaking step can be prolonged to an overnight incubation at 4°C.

3. 10. Cell culture and immunofluorescence staining

Direct comparison between direct immunohistochemistry and immunofluorescence is shown in figure 3-15. In our experimental setting, human neuroblastoma cells SK-N-SH were maintained in DMEM medium supplied with

10% fetal bovine serum (FBS). 25×10^3 cells were seeded onto an 8 wells Nunc™ Lab Tek™ II chamber and were incubated for 24h and 48h with 20%(w/v) tissue homogenates from 97Q, 24Q and Wt animals. After 24h and 48h, the cells were washed with cold PBS and prepared for membrane staining. Cells were fixed with 4% formaldehyde for 20 min at room temperature. Wheat Germ Agglutinin conjugates solution were prepared in PBS (invitrogen®). $10 \mu\text{g}/\text{ml}$ of solution was used in each well and incubated for 10 min at room temperature. The cells were washed twice and permeabilized with 0.1% Triton-100 for 5 min. To block the unspecific binding of the first antibody, the cells were incubated with 3%FCS in PBS for one hour and incubated with primary anti-AR antibody for one hour. Finally, the cells have been incubated with secondary fluorescent mouse antibody Alexa-Flour-488. A confocal Leica microscope was used to capture images. Image analysis was completed with the same microscope and image J software. From each sample, 100 cells were counted and used for quantification analysis of presence of nuclear staining or cytoplasmic staining.

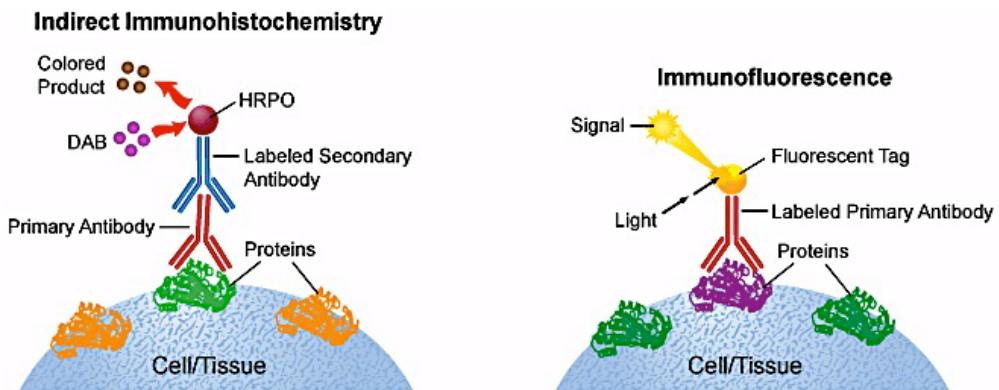
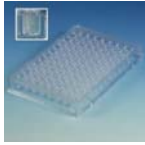


Figure 3-15: Schematic comparisons between direct immunohistochemistry and immunofluorescence. Both methods have been used in this thesis. Brief description of immunohistochemistry comes in the following chapter (Adapted from Leinco® Technologies, Inc).

3.10.1 Cellular uptake and cell toxicity studies

Human neuroblastoma cells SK-N-SH were maintained in DMEM medium supplied with 10% fetal bovine serum (FBS). 5×10^3 cells were seeded onto a 96 wells plate and were incubated for 24h and 48h with 20%(w/v) tissue homogenates from 97Q, 24Q and Wt animals (Fig. 3-16). 20%w/v total tissue homogenates from transgenic animals of 97Q and 24Q were prepared as explained before. Wt mouse tissues with identical age were also prepared. CellTiter®-Blue cell viability assay was done in 96-well plates. $20 \mu\text{l}$ of CellTiter®-Blue Reagent is added directly to each as well, the plates were incubated at 37°C to allow cells to convert resazurin to resorufin and after shaking the plate for 10 sec, the fluorescence signal was measured at $560_{\text{Ex}}/590_{\text{Em}}$ nm. The average of fluorescence values of the culture medium background was subtracted from all experimental wells.

5000 SK-n-SH cells



after 24h

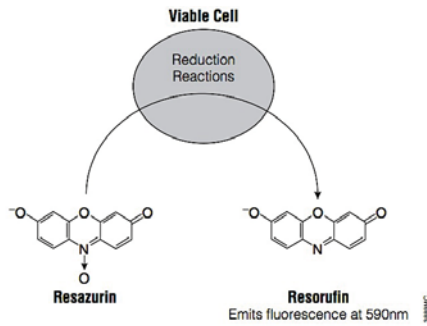


Figure 1. Conversion of resazurin to resorufin by metabolically active cells results in the generation of a fluorescent product. The fluorescence produced is proportional to the number of viable cells.

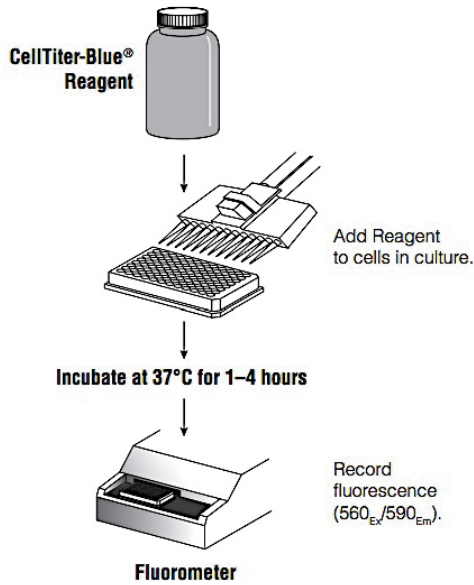


Figure 3-16. Scheme of CellTiterBlue cell viability assay. The visible light absorbance properties of this reagent undergo a blue shift upon reduction of resazurin to resorufin. Fluorescence was used to record the results, as it is more sensitive (Adapted from CellTiterBlue® web page).

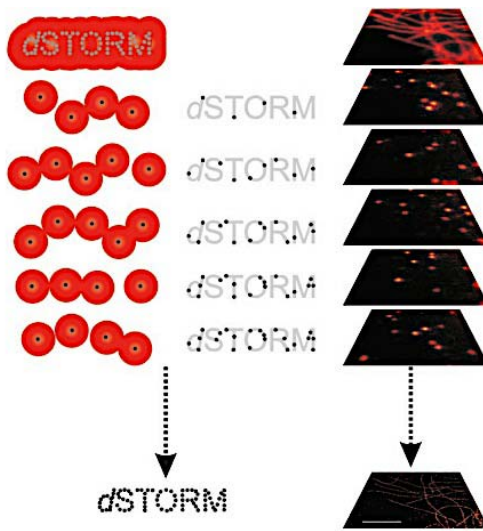


Figure 3-18: The dSTORM concept. The target molecule of interest is labeled with photoswitchable fluorophores. The top image on the right side shows labeled microtubules in a mammalian cell. At the beginning of the experiment, all fluorophores are transferred to the non-fluorescent OFF state upon irradiation with light of appropriate wavelength and intensity. Either spontaneously or photo-induced upon irradiation with a second laser wavelength, a sparse subset of fluorophores is reactivated. If the probability of activation is sufficiently low, then the activated fluorophores

residing in their ON state are statistically spaced further apart than the resolution limit and their positions can be precisely determined. Repetitive activation, localization and deactivation allow a temporal separation of spatially unresolved structures in a reconstructed image (image below). Scale bar, 2 μm (Biozentrum Universität Würzburg) (202).

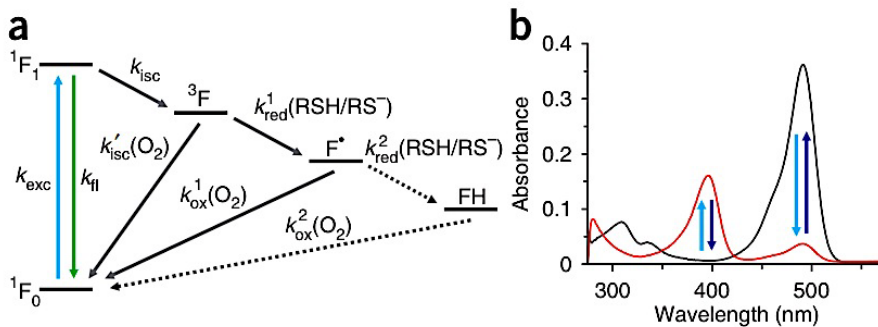


Figure 3-19: Photoswitchable dyes are used for dSTORM. Reversible photoswitching of Alexa Fluor and ATTO dyes in the presence of thiols. (a) The fluorophore is either cycled between its singlet ground and excited state emitting fluorescence photons or can undergo intersystem crossing with rate k_{isc} upon irradiation. The triplet state ($3F$) can react with molecular oxygen to recover the singlet ground state and produce singlet oxygen or react with the thiolate with rate $k_{red}^1(RSH/RS^-)$ to form the radical anion of the fluorophore (F^\bullet) and the corresponding thiyl radical. The radical anion can be oxidized by oxygen with rate $k_{ox}^1(O_2)$ to recover the singlet ground state. Because radical anions of most rhodamine and oxazine derivatives show pronounced absorption at ~ 400 nm irradiation, for example, at 405 nm promotes recovery of the fluorescent form (i.e., OFF- and ON-switching are photoinduced). Whereas the thiyl radicals formed react efficiently with molecular oxygen to produce superoxide radicals and hydrogen peroxides, the fluorophore radical anion is very unreactive and survives for up to several seconds even in the presence of molecular oxygen. Fluorophores such as ATTO 655 and ATTO 680 accept a second electron to the fully reduced leuco-form (FH). Oxidation of FH with oxygen also recovers the ON state. (b) Absorption spectra of Alexa Fluor 488 in PBS (pH 9.3) in the presence of 100 mM MEA. After irradiation at 488 nm (light blue), the absorption at 488 nm decreases and the radical anion absorbing maximally around 400 nm appears. The fluorescent state is recovered spontaneously or by direct excitation of the radical anion at ~ 400 nm (dark blue). The stability (lifetime) of the radical anion is mainly determined by the oxygen concentration and can easily exceed several hours (Biozentrum Universität Würzburg) (202, 203).

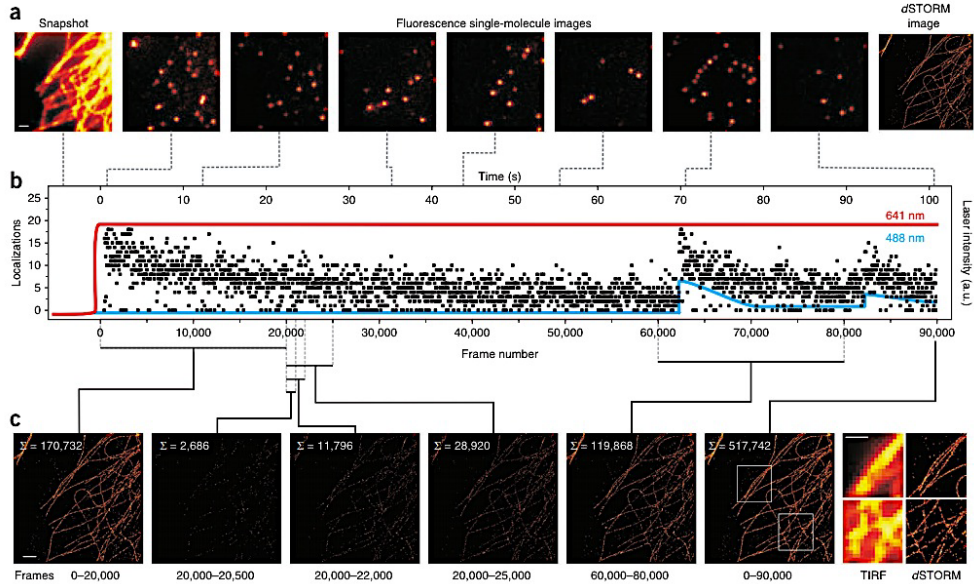


Figure 3-20: dSTORM acquisition procedure exemplified for Alexa Fluor 647-labeled β -tubulin in a COS-7 cell. (a,b) A fluorescence image of the structure is measured at low excitation intensity (a) ($< 0.1 \text{ kW cm}^{-2}$ see red line in b). In the next step, higher irradiation intensity is applied to transfer the majority of fluorophores into a non-fluorescent OFF state until a sufficiently low spot density is reached. Finally, a super-resolved image is reconstructed from all localizations. (b) Number of localizations per image frame (approximately ten, corresponding to ~ 0.1 spots per μm^2) plotted against the frame number and time, respectively. The images were acquired with a frame rate of 885 Hz at an excitation intensity of 30 kW cm^{-2} at 641 nm and additional irradiation at 488 nm (blue line) with $0\text{--}0.5 \text{ kW cm}^{-2}$. Direct excitation of the OFF state at 488 nm increases the number of fluorophores residing in the ON state and has to be carefully controlled. (c) Image reconstruction showing that a highly resolved image can already be reconstructed from 20,000 images corresponding to a total acquisition time of ~ 20 s. The structure cannot be fully resolved after analysis of $< 2,000$ frames. Σ , the number of localizations used to reconstruct the dSTORM image. Scale bars, $1 \mu\text{m}$ (202).

3.11.2 STED (Stimulated emission depletion microscopy)

STED microscopy shares principles of confocal laser scanning microscopy and structured illumination microscopy. In traditional confocal microscopy, a laser beam is focused to diffraction-limited spot, referred to as the point spread function, PSF. The PSF has a diameter of ca 250 nm and excites fluorescence in the sample, which is collected as the beam is raster-scanned across the region of interest (Fig. 3-23). One can think of the laser beam as a paintbrush - the thinner the tip, the finer the detail we can recover from the sample ('Excitation beam' shown in blue in figure below). The idea behind STED, or more generally the RESOLFT principle, is to 'shape' the tip into a spot much smaller than the 250 nm diffraction limit, by using a so-called depletion beam, which has a doughnut shaped intensity distribution, with zero intensity in the middle and high intensity in an annular region around the central position ('STED beam' in Fig. 3-21). Superimposing this depletion beam on the excitation beam prevents spontaneous emission (fluorescence) from fluorophores anywhere except in a

very small central region, thus yielding an effective PSF of only 30 to 90 nm in diameter (Fig. 3-21 and 3-22). This permits one to gain information on a much smaller scale than possible with standard confocal imaging. The technique works at a reasonable depth and is thus useful to look at distribution of molecules deep within cells.

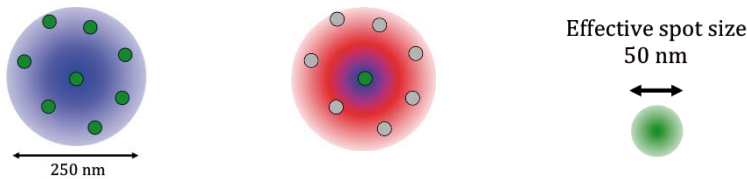


Figure 3-21: STED is a confocal technique, which uses a doughnut shaped laser beam to ‘switch off’ fluorescence outside of the centre of the beam (Laser analytics group, Cambridge).

STED principles

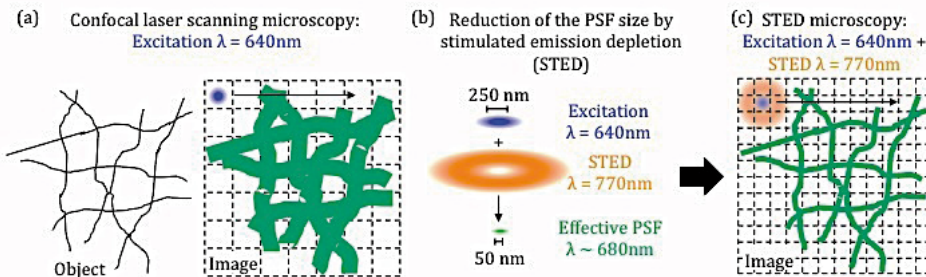


Figure 3-22: Principle of Stimulated Emission Depletion microscopy (STED)(Adapted from Pierre Mahou, Cambridge)

The excitation of various fluorophores is enabled by a super-continuum source generated by pumping a photonic crystal fiber and the STED beams is produced by a spatial light modulator that also corrects for the optical aberrations of the system. This microscope has been successfully evaluated with ideal test samples (Fig. 3-23 a-b) and is now being used to monitor the formation of amyloid fibrils *in vitro* (Fig. c). In this thesis, 20%w/v total muscle tissue homogenate of 97Q and 24Q mice were dried and fixed on a Lab-Tek™ chamber. Using specific primary antibody for elongated version of polyQ tract (1C2) and STED compatible secondary antibody (ATTO 647 anti-mouse antibody), the aggregates in mouse muscle total tissue homogenate have been stained and imaged using 647nm Laser.

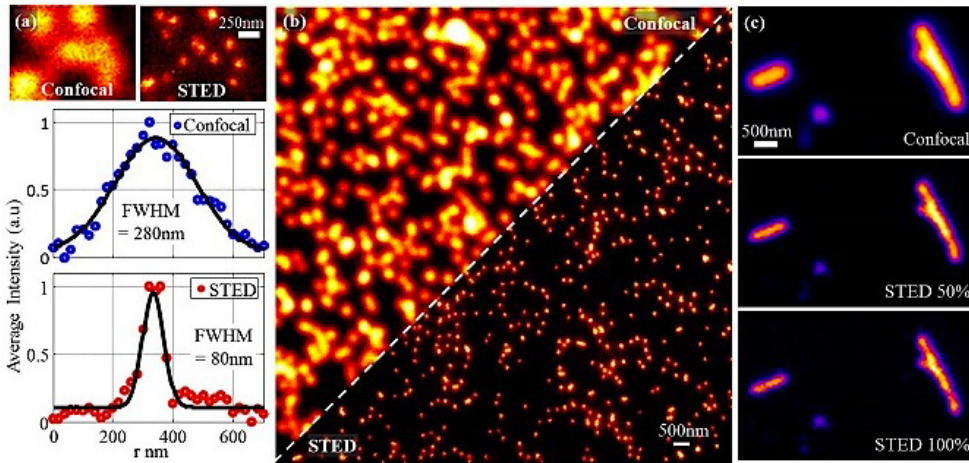


Figure 3-23. Evaluation of the performance of the lab-built STED microscope from ideal calibration samples. (a) Estimation of the resolution of the confocal/STED modality from 40nm beads. The resolution enhancement is on the order of 3.5. (b) Performance of the STED modality in a case of a dense sample formed by 100nm beads. (c) Confocal/STED images of amyloid fibrils labeled with the dye STAR 635P (Laser analytics group, Cambridge).

3.12. Atomic force microscopy (AFM)

Aggregates were captured by Seprion beads and eluted with 50 μ l 0.75%SDS. From each preparation 5 μ l was deposited on freshly cleaved mica plate (Ted Pella Inc, Redding, CA) and incubated overnight at RT to get dry. The captured material was washed with 200 ml of ultrapure water and dried under a gentle stream of air. Aggregates from tissue homogenates were prepared by using Seprion beads as described previously. Each deposition was imaged using a Bruker Dimension FastScan[®] AFM microscope. Images were taken with silicon cantilevers with a nominal spring constant of 40 N/m and resonance frequency of approximately 300 kHz (Fig. 3-24). Typical imaging parameters were: drive amplitude 150–500 mV with set points of 0.7–0.8 V, scan frequencies of 2–4 Hz, image resolution 512x512 points, and scan size of 3 μ m. All the experiments were performed in 5 different tissues from each group of animals and with many replicas for each quantitative data. Several images were obtained from separate locations across the mica surfaces to ensure reproducibility. For quantitative analysis of AFM images, size analysis of aggregates observed by AFM was performed and images were analyzed with the image processing toolbox program. Individual aggregates in an AFM image are automatically located and their volumes and heights and other geometrical characteristics are measured, facilitating a quick analysis of thousands of individual aggregates.

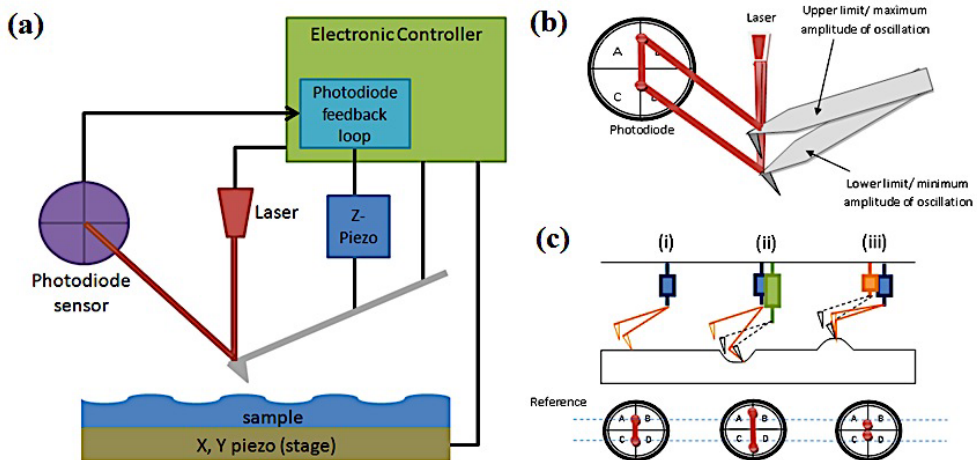


Figure 3-24: (a) Schematic description of an AFM. (b) Cantilever deflection and laser spot on the photodiode, (c) schematic of the tapping mode operation. Adapted from (204).

3.13. Transmission electron microscopy (TEM)

The TEM operates on the same basic principles as the light microscope (Fig. 3-25) but uses electrons instead of light. Because the wavelength of electrons is much smaller than that of light, the optimal resolution attainable for TEM images is in many orders of magnitude better than that from a light microscope. Thus, TEMs can reveal the finest details of internal structure - in some cases as small as individual atoms. Transmission electron microscopy uses high-energy electrons (up to 300 kV accelerating voltage) that are accelerated to nearly the speed of light. The electron beam behaves like a wavefront with wavelength about a million times shorter than light waves. When an electron beam passes through a thin-section specimen of a material, electrons are scattered. A sophisticated system of electromagnetic lenses focuses the scattered electrons on an image or a diffraction pattern, or a nano-analytical spectrum, depending on the mode of operation.

Each of these modes offers a different insight about the specimen. The imaging mode provides a highly magnified view of the micro- and nanostructure and ultimately, in the high resolution imaging mode a direct map of atomic arrangements can be obtained (high resolution EM = HREM). The diffraction mode (electron diffraction) displays accurate information about the local crystal structure. The nanoanalytical modes (x-ray and electron spectrometry) tell researchers which elements are present in the tiny volume of material. These modes of operation provide valuable information for scientists and engineers in search of stronger materials, faster microchips, or smaller nanocrystals.

In our experiment, for immune labeling, 20 μ l eluted material from the Seprion bead capture assay was dried at 95 $^{\circ}$ C in a heating block and resuspended in 5 μ l

ultrapure water. Three microliters were transferred to a freshly glow-discharged Formvar/carbon-coated grid and incubated at RT for 1 min. Excess solution was removed and the grid allowed to air-dry for 5 min. Grids were rinsed briefly in PBS and transferred to blocking solution (0.1% BSA-CTM (Aurion) in PBS) for 15 min. This was followed by incubation with H280 primary antibody (1:200) diluted in blocking solution for one hour.

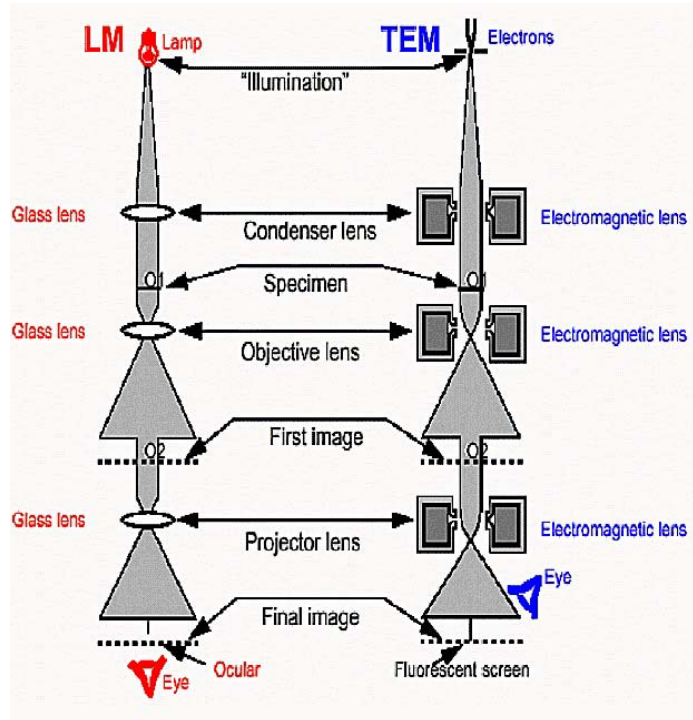


Figure 3-25: Schematic comparison between transmission electron microscope (TEM) and light microscope (LM) (Adapted from Martin Saunders, CMM).

The grids were washed six to eight times in a drop of blocking solution (2 min/drop) and transferred to Rabbit secondary antibody conjugated with 10 nm colloidal gold particles (BB International) diluted 1:200 in blocking solution for 1 h. The grids were washed as above, followed by two rinses in a drop of ultrapure water (1 min/drop) and air-dried for 3 min. False positive or negative staining were performed by adding 3 μ l of 1% uranyl acetate for 45 s, excess stain was removed with hardened filter paper and grids allowed to air-dry or the procedure terminate with a quick ultrapure water wash. Images were taken with a transmission electron microscope at 120 kV.

3.14. Immunohistochemical analysis

Immunohistochemistry (IHC) combines histological, immunological and biochemical techniques for the identification of specific tissue components by means of a specific antigen/antibody reaction tagged with a visible label. IHC

makes it possible to visualize the distribution and localization of specific cellular components within a cell or tissue. In contrast to western blotting and ELISA, immunohistochemistry (IHC) allows the study of the distribution and localization of specific cellular components in cells or tissues. However, IHC still uses labeled antibody to trace antigen-antibody interactions and the antibody can be detected by several methods. Most commonly, a secondary antibody conjugated to an enzyme (horseradish peroxidase or alkaline phosphatase) is used; this binds to the primary antibody and indirectly stains the antigen of interest. A similar technique involving fluorophore-conjugated secondary antibodies is called immunofluorescence. Here, some major aspects of IHC, including sample preparation, antigen retrieval, as well as schematic steps in immunohistochemical staining were explained (Fig. 3-26).

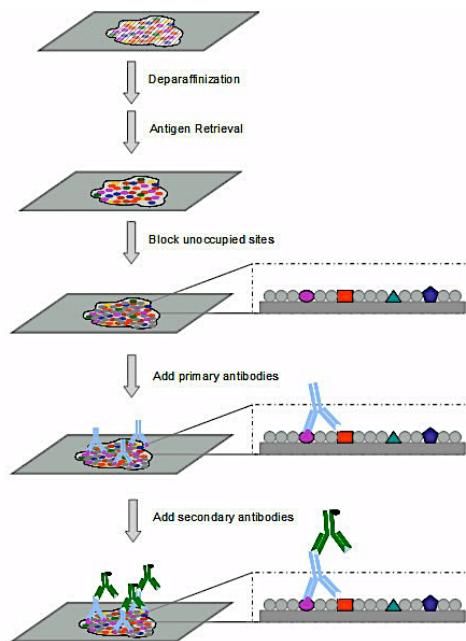


Figure 3-26: Schematic steps in immunohistochemical staining (Adapted from Proteintech® web page).

1. Sample preparation: Prompt and adequate tissue preparation is crucial in immunohistochemistry. Although there is no single universal method for tissue fixation, many antigens can be successfully detected in formalin-fixed paraffin-embedded tissue sections. Alternatively, sections may be prepared by snap freezing in liquid nitrogen and sectioned with a cryostat, followed by fixation with cold acetone or alcohol. However, there are several disadvantages associated with frozen sections, including poor morphology, poor resolution at higher magnifications, the need for special storage, only limited retrospective studies possible and the current cutting difficulty. Paraffin-embedded sections overcome most of these problems and are the most widely used preparation method. This is the same approach we used in our experiments.

2. Antigen retrieval Pre-treatment with the antigen retrieval reagent is needed to break the protein cross-links formed by formalin fixation in order to uncover hidden antigenic sites. The methods used include the application of heat (HIER: Heat Induced Epitope Retrieval), use of digestive enzymes (PIER: Proteolytic Induced Epitope Retrieval) or a combination of the two. While heat-mediated antigen retrieval is commonly performed by a microwave, a pressure cooker or an autoclave, enzyme-mediated antigen retrieval uses a combination of digestive enzymes such as proteinase K, trypsin, chymotrypsin, pepsin and pronase. However, the use of enzyme digestion could inconsistently destroy some epitopes and tissue morphology, leading to false negative results.

In our experimental condition, we perfused 20 ml of a 4% paraformaldehyde fixative in phosphate buffer (pH 7.4) through the left cardiac ventricle of mice deeply anesthetized with ketamine-xylazine, postfixed tissues overnight in 10% phosphate-buffered formalin, and processed tissues for paraffin embedding. Then we deparaffinized 4 μ m thick tissue sections dehydrated with alcohol, treated in formic acid for 5 min at room temperature. The tissue sections were blocked with normal horse serum (1: 20) and incubated with mouse anti-expanded polyQ, 1C2 (1: 20 000, Chemicon®, Temecula, CA, USA). The sections were then incubated with biotinylated anti-mouse IgG (1: 1000, Vector Laboratories, Burlingame, CA, USA). Immune complexes were visualized using streptavidin–horseradish peroxidase (Dako®, Glostrup, Denmark) and 3,3'-diaminobenzidine (Dojindo, Kumamoto, Japan) as described before (163).

3.15. Synchrotron micro-computed tomography

During the last decade, many progresses in the field of tomography have been made. Scientist used the concept of tomography to go beyond histology. Micro-tomography started to be used as a methodology for the biological tissue characterization.

Current light microscopy methods such as serial sectioning, confocal microscopy or multi-photon microscopy are severely limited in their ability to analyze rather opaque biological structures in three dimensions, while electron optical methods offer either a good three-dimensional topographic visualization (scanning electron microscopy) or high resolution imaging of very thin samples (transmission electron microscopy). However sample preparation commonly results in a significant alteration and the destruction of the three-dimensional integrity of the specimen. Depending on the selected photon energy, the interaction between X-rays and biological matter provides semi-transparency of the specimen, allowing penetration of even large specimens. Based on the projection-slice theorem, angular projections can be used for tomographic imaging. This method is well developed in medical and materials science for structure sizes down to several micrometers and is considered as being non-destructive. Achieving a spatial and structural resolution that is sufficient for the imaging of cells inside biological tissues is difficult due to several experimental conditions. Major problems that

cannot be resolved with conventional X-ray sources are the low differences in density and absorption contrast of cells and the surrounding tissue. Therefore, X-ray monochromatization coupled with a sufficiently high photon flux and coherent beam properties are key requirements and currently only possible with synchrotron-produced X-rays. In this study, we report on the three-dimensional morphological characterization of muscle tissues and neuromuscular junctions using synchrotron-generated X-rays demonstrating the spatial distribution of single cells inside the tissue, single cell degeneration and their quantifications, while comparing our findings to conventional histological techniques.

In this thesis, we used this approach to see the structural differences between different muscle tissues of Wt and transgenic mice of SBMA. Hind limb muscle tissues and spinal cord have been prepared in 2-3 mm in diameter and 5-6mm in length and fixed in 4% Para-formaldehyde (PFA) for minimum of 24h and aligned in agarose gel. The samples were introduced to synchrotron. The analyses have been done in Advanced Photon source (APS) operated by the University of Chicago. This work was done in collaboration with the Karlsruhe Institute of Technology (KIT) and KIT managed all the data analysis.

3.16. Statistical analysis

Statistical tests were performed using the GraphPad Prism program. For column statistics, one-way ANOVA with Bonferroni or Tukey's multiple comparison tests was performed. $P \leq 0.05$ was considered significant.

PART IV.

Chapter 4. Sequence determinants of AR oligomerization in SBMA

4.1. Introduction

4.2. Results

4.3. Summary

Chapter 4. Sequence determinants of AR oligomerization in SBMA

4.1. Introduction

Spinal bulbar muscular atrophy (SBMA) is a rare neuromuscular disease, with a prevalence of one per 50,000 males that belongs to a group of nine hereditary disorders termed polyglutamine (*polyQ*) diseases. These occur as a result of the expansion of polymorphic polyQ tract in proteins that are otherwise unrelated (19). The unstable nature of such tracts is due to the propensity of CAG and GTC codons repeats that codify for them and that form non-B-DNA structures, such as hairpins that cause DNA slippage during replication (205, 206) As a consequence of this expansion, the codified proteins are often found to misfold, oligomerize and aggregate to form fibrillar species resembling amyloid fibrils (207–209).

The polyQ tract responsible for SBMA is found in the N-terminal transactivation domain (NTD) of AR. AR is a nuclear receptor activated by androgens such as dihydrotestosterone (DHT) that regulates the expression of the male phenotype (Fig. 4-1A). Although the molecular basis of SBMA is not yet fully established, there are strong indications that AR aggregation is the main cause of the disease. In fact, this hypothesis is supported by the observation that SBMA patients exhibit AR loss of function symptoms attributed to the depletion of the soluble receptor, such as gynecomastia and infertility, as well as a gain of function symptoms, such as progressive muscle cramps, fasciculations and dysphagia (210), caused by the specific toxicity of AR aggregates to muscle cells and motor neurons (22, 207, 211–214).

Prior to activation by androgens, AR is a cytosolic protein in complex with molecular chaperones that bind to the NTD (residues 1-558), which is intrinsically disordered (215), and stabilize it against degradation and aggregation. Androgen binding to the ligand-binding domain (LBD) (residues 670-919, Fig. 4-1A) leads to a collective conformational change that causes the disassembly of this complex and targets the receptor to the nucleus by exposure of a nuclear localization signal found in the hinge region that connects the LBD with the DNA-binding domain (559–631, Fig. 4-1A). The DBD of nuclear AR interacts with specific DNA sequences known as androgen response elements (AREs) found in promoters and enhancers of genes regulated by AR. Once AR is bound to DNA, the NTD recruits the general transcription factors and transcriptional co-regulators needed to assemble the transcription machinery, leading to transactivation.

The polymorphic polyQ tract in AR starts at position 58 and can be between 14 to 34 residues long in healthy individuals. Sizes over 37 residues are associated with SBMA and the length of the polyQ tract anti-correlates with the age of onset (216). Although the function of the polyQ tract is not yet known there are

indications that it plays a role in the regulation of AR function as individuals with short polyQ tracts exhibit a higher propensity to suffer from prostate cancer, a condition that relies on AR expression and activation for its progression (217–219).

There is evidence from *Drosophila* and mice models that the development of the SBMA phenotype requires AR activation (84, 220). These observations are in agreement with a pathogenesis mechanism in which AR aggregation occurs when its complex with molecular chaperones is dissociated after androgen binding. Given that activation leads to nuclear translocation, this hypothesis is also in agreement with the nuclear localization of AR aggregates, known as nuclear inclusions (NIs). Finally, additional support results from studies of the effect of AR antagonists on mice expressing human AR with an elongated polyQ tract; whereas untreated mice have a phenotype closely resembling that of SBMA patients, those treated with AR antagonists appear healthy (94, 163, 221, 222).

Although the polyQ tract of AR is likely to play an important role in the early stages of AR aggregation in SBMA, neither the structure of the tract nor the oligomerization mechanism of the receptor have been characterized. This acts as an important hurdle for the identification of therapeutic approaches based on the inhibition of AR aggregation by small molecules or biologics. To remedy this, we have used NMR in solution to study the structural properties of an N-terminal fragment of AR, found in AR aggregates linked to SBMA (223, 224), that is the product of proteolytic cleavage by caspase 3 (residues 1-153, Figs. 4-1A and 4-1B) and whose aggregation is thought to play a key role in the onset of the disease.

Despite the presence of a large number of consecutive amino acids of the same type, the NMR strategy that we have used enabled us to characterize the structural and dynamical properties of every residue. Our findings indicate that although the protein is intrinsically disordered the polyQ tract displays substantial helical propensity, which increases with the length of the tract to reach values compatible with the formation of a helix. We also observe that the oligomerization rate of the protein does not greatly depend on the length of the polyQ tract because it is driven by inter-molecular interactions involving a region of sequence, with sequence $^{22}\text{GAFQNLFQSVREVIQ}^{36}$, that is separated from it by 20 residues (Fig. 4-1B).

Our results provide a striking example on how the regions flanking the polyQ tract can influence both the structural properties of the tract and its mechanism of aggregation. In addition they identify the region of sequence that nucleates the oligomerization of AR in SBMA, the region $^{22}\text{GAFQNLFQSVREVIQ}^{36}$, as a therapeutic target for this currently incurable and relatively neglected rare disease.

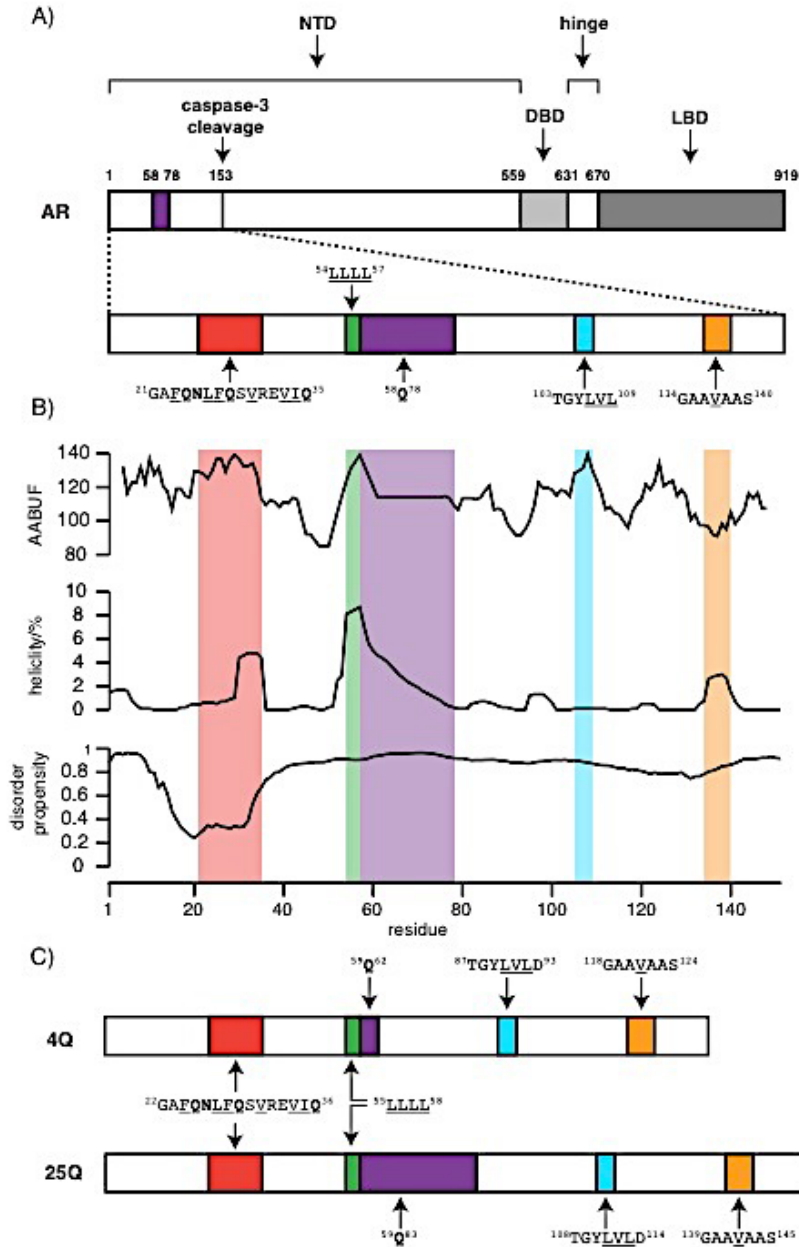


Figure 4-1: A) Domain organization of AR with an indication of the position of the polymorphic polyQ tract in purple, of the Caspase 3 cleavage site and, in the resulting fragment, of the $^{22}\text{GAFQNLFSVREVIQ}^{35}$, $^{55}\text{LLLL}^{57}$, $^{105}\text{TGYLVL}^{109}$ and $^{134}\text{GAAVAAS}^{142}$ motifs, which are colored red, green, blue and orange, respectively; in this and the following panel the residues are numbered according to the Uniprot entry for AR (www.uniprot.org/uniprot/P10275) B) AABUF, predicted helicity and disorder propensity of the fragment reported as a function of residue number. C) 4Q and 25Q constructs used in this work, where the difference in the numbering with panels A and B, is due to the addition of a residue at the N-terminus after cleavage of the MBP by using the TEV protease. In the motifs highlighted in the figure, hydrophobic residues are underlined and Q and N residues are shown in bold.

4.2. Results

4.2.1 Preparation of solutions of monomeric AR fragments

Preparing solutions of monomeric aggregation-prone proteins is challenging due to the propensity of these species to form aggregates, but essential for characterizing reproducibly and reliably their structural properties (225, 226). We expressed two AR constructs containing a polyQ tract of 4 and 25 Gln residues (4Q and 25Q- Fig. 4-1C), fused to His₆-tagged maltose binding protein (MBP) in *E. coli*. After purification by affinity and size exclusion chromatography and cleavage of the fusion protein by the tobacco etch virus (TEV) protease, the fragments were rendered monomeric by applying a stringent disaggregation procedure based on the one recently put forward for the preparation of monomeric A β ₁₋₄₂ solutions (226).

After the disaggregation procedure, solutions of 4Q and 25Q were proven to be more than 98% monomeric according to the relative volume distribution of dynamic light scattering (DLS) data (Fig. 4-2A). The hydrodynamic radius of 4Q and 25Q calculated from the DLS data was *ca* 3.5 nm, which is in good agreement with the values determined from the elution volume obtained by size exclusion chromatography (SEC). The oligomeric state of the samples was also analyzed by sedimentation velocity analytical ultracentrifugation (SV-AUC) analysis, which yielded a narrow distribution of particles centered on a main species with a sedimentation coefficient of 1.1 S (for 4Q) and 1.3 S (for 25Q). Data fitting performed by the software SEDFIT (227) revealed a low but detectable population of a second species (1-3%), with a sedimentation coefficient of 2.6 S (4Q) and 2.8 S (25Q). By comparison of the experimental frictional ratios with their theoretical counterparts for different oligomers we obtained that, for both constructs, the major species corresponded to the monomer whereas the minor species likely corresponded to a tetramer (Fig. 4-2B). A similar distribution of oligomeric states was recently described for synthetic peptides corresponding to exon 1 of the Htt that also harbors a polyQ tract (22).

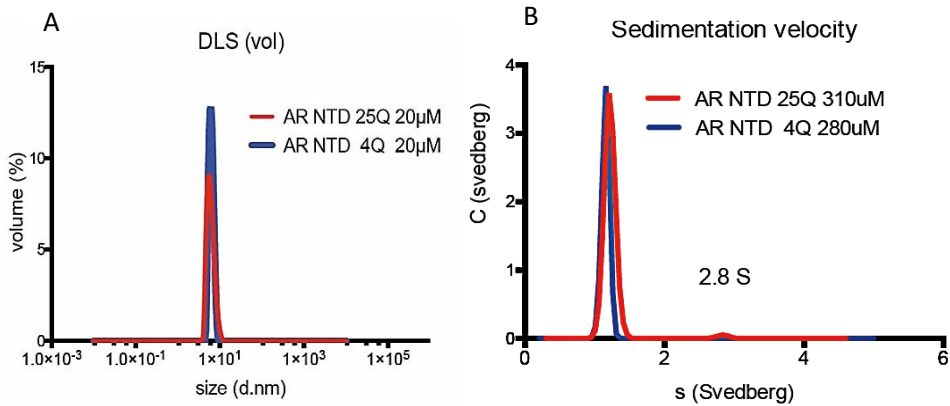


Figure 4-2: A) Dynamic light scattering size distribution by volume of 20µM 4Q (blue) and 25Q (red) immediately after preparation at 37°C. The mean size value is 5.6 nm and 6.1 nm respectively. B) Sedimentation velocity analytical ultracentrifugation of 4Q (blue) and 25Q (red) of stock solutions of respectively 310 µM and 280µM protein immediately after preparation.

4.2.2 Analysis of AR oligomerization

To characterize the oligomerization process of 4Q and 25Q, we monitored the time evolution of 20 mM samples of these constructs by time resolved dynamic light scattering (DLS), a technique that is very sensitive to the presence of soluble oligomers. An analysis of the weighted mean diameter of the ensemble of particles (Z-average) revealed that oligomerization occurs in solutions both of 4Q and 25Q (Fig. 4-3A). These results indicated that oligomerization occurs independently of the length of the polyQ tract and is therefore likely to be triggered by a different region of sequence.

To characterize the structural properties of the monomeric proteins and their evolution during oligomerization, we used circular dichroism (CD) and nuclear magnetic resonance (NMR) spectroscopies. CD spectra collected immediately after preparing the monomeric protein solutions indicated that the constructs were substantially disordered, but with a fractional content of α -helix, which was higher for 25Q than for 4Q (Fig. 4-3C). These measurements, in agreement with previous studies (228), suggest that the polyQ tract of AR is partially helical. To study how the secondary structure evolved with time, a solution of 20 µM 25Q was analyzed both by CD and DLS at various time points within 150 hours of incubation. The CD spectra showed no apparent change in secondary structure as oligomerization progressed which indicates that the soluble oligomers are essentially disordered or that they represent a small fraction of the sample and therefore contribute slightly to the CD signal (Fig. 4-3B).

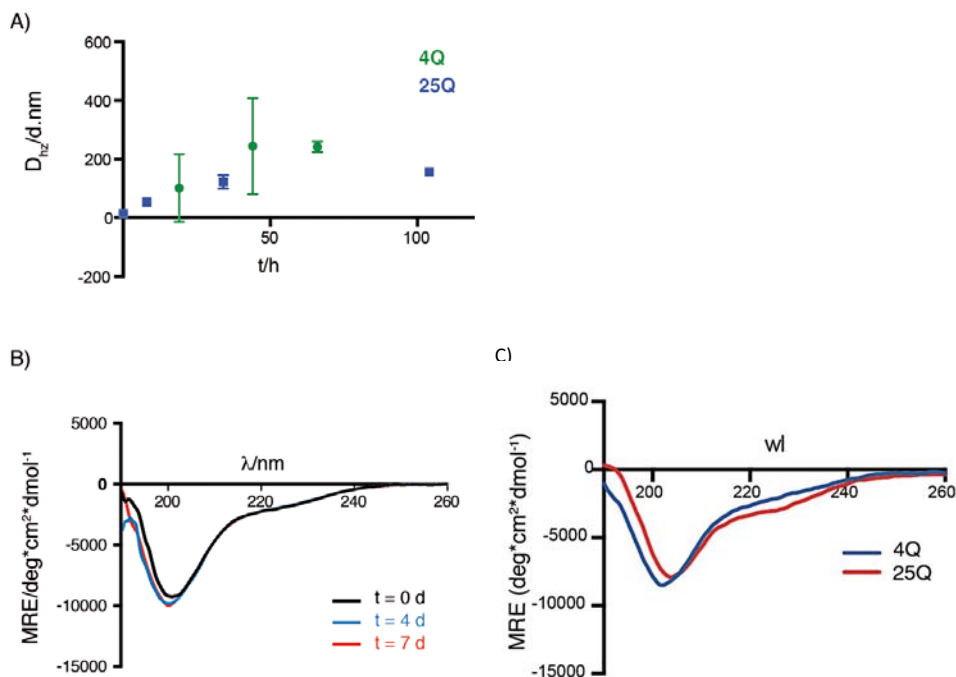


Figure 4-3: A) Time resolved analysis of 20 μ M solutions of 4Q and 25Q oligomerization by DLS. B) CD spectrum of 20 μ M SDP solution 25Q after 0 h, 96 h and 170 h of incubation at 37°C C). CD spectrum of 130 μ M solutions 4Q and 25Q immediately after sample preparation.

Time-resolved NMR experiments were performed in an attempt to identify the regions of sequence involved in the inter-molecular interactions that stabilize the oligomers. 15 N-labeled 50 μ M samples of 25Q were incubated at the same conditions used for the DLS and CD measurements and 1 H- 15 N HSQC experiments were acquired at 18.8 T (800 MHz) at various time points. A fraction of the cross-peaks were observed to evolve upon oligomerization mainly by decreasing in intensity but also, in some cases, by splitting in two or, in one case, in three (Fig. 4-4). That only a fraction of the resonances evolve with time indicates that the changes in chemical environment that occur as a consequence of oligomerization are specific to certain residues.

In some instances the time-evolution of the resonances allowed observing the build-up and decay of signals with well-defined chemical shifts, as illustrated in Figure. 4-9, indicating that oligomerization occurs with the formation of specific on-pathway intermediates. The time-evolution of 25Q samples under these conditions, due to oligomerization, precluded assignment of the resonances by using multidimensional NMR and it therefore became necessary to identify a sample preparation procedure that led to samples that were sufficiently kinetically stable for this purpose.

4.2.3 Preparation and structural characterization of a kinetically stable sample by NMR

In an attempt to produce kinetically stable samples, a 400mM solution of 25Q was incubated in 6 M urea overnight and then dialyzed (1:1000) against 20 mM sodium phosphate buffer pH 7.4. The ^1H - ^{15}N HSQC spectrum of the sample obtained by this procedure (urea incubation followed by dialysis, UID) was very similar to that of the 25Q obtained by SDP and did not change significantly over the period of time required to acquire the multidimensional NMR experiments needed for the assignment of all the resonances. In addition, DLS analyses confirmed that the samples were kinetically stable.

Sequence-specific assignments for both UID 25Q and 4Q were obtained by acquiring a set of 4D ^{13}C -detected NMR experiments selected among those proposed by Bermel *et al.* 2012 (133, 229). Heteronuclei were exploited to take advantage of their improved chemical shift dispersion relative to that of protons and their reduced sensitivity to exchange processes. The problem of extensive cross-peak overlap in the spectra was overcome by acquiring high-dimensional experiments, exploiting non-uniform sampling (NUS) to reduce the experimental time while preserving high spectral resolution in the indirect dimensions. In addition, two 4D TROSY $^1\text{H}^{\text{N}}$ -detected experiments, variants of those proposed by (230, 231), were recorded to confirm the assignment obtained with the ^{13}C -detected experiments. The assignments of H^{N} , H^{α} , H^{β} , C' , C^{α} , C^{β} and N resonances of 4Q and 25Q are reported in the BMRB (<http://www.bmrb.wisc.edu/>).

The complete 2D ^1H - ^{15}N HSQC (Fig. 4-4) and 2D CON-IPAP (Fig. 4-5A) spectra of 4Q and 25Q and the central regions of both spectra are shown in Figure. 4-5B, which illustrates that the spectral differences caused by the elongation of the polyQ tract are limited to the resonances of the tract itself and to the regions of sequence immediately adjacent it, particularly at its N-terminus. The assignment of the polyQ tract of 25Q appeared as particularly challenging due to the expected extensive overlap of the glutamine resonances in all spectra. Interestingly, however, both in the 2D ^1H - ^{15}N HSQC and 2D CON-IPAP spectra, the resonances of the tract were arranged along a pseudo-diagonal in which the cross-peaks appear as ordered according to the position of the corresponding residue in the sequence.

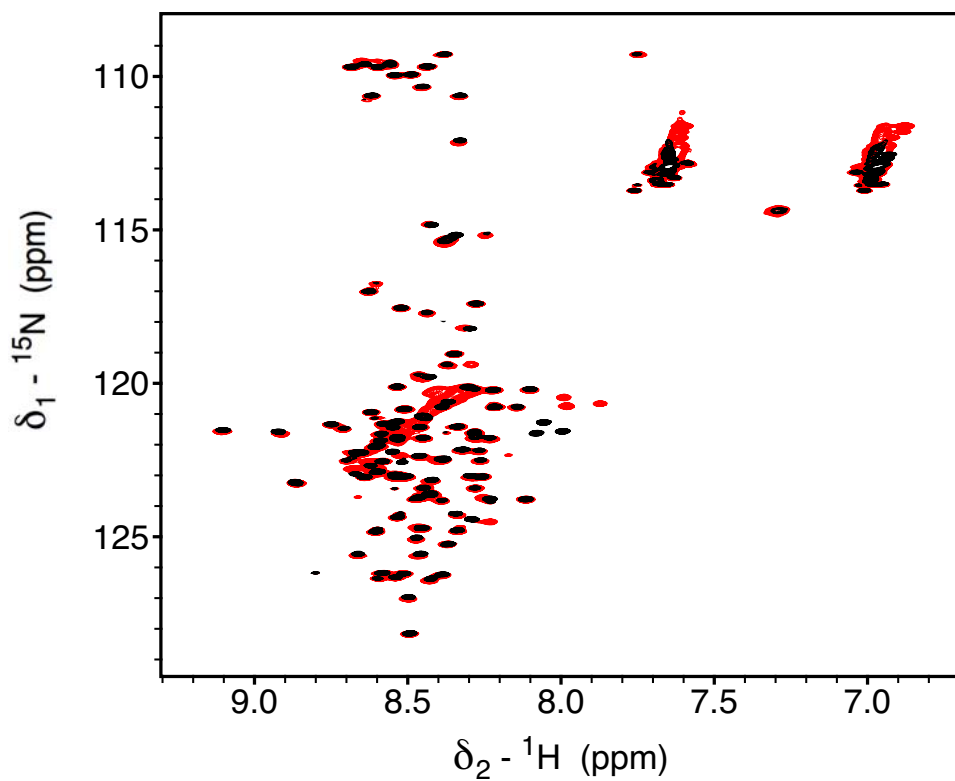


Figure 4-4: The complete ^1H - ^{15}N HSQC spectrum and with an indication of the resonances that experiences the largest shifts upon increasing the length of the polyQ tract.

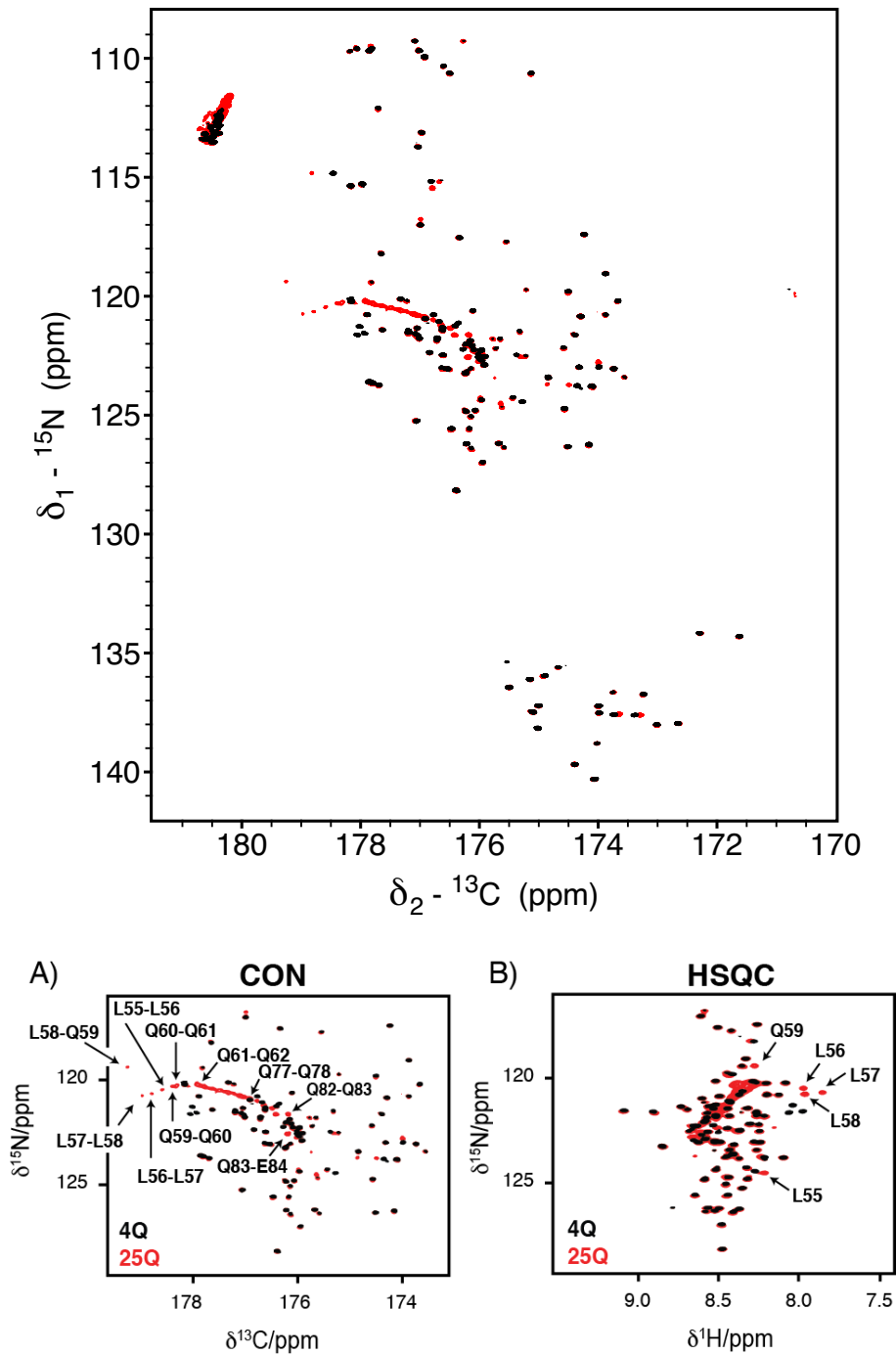


Figure 4-5: A) the CON-IPAP spectrum of 4Q (black peaks) and 25Q (red peaks) B) Central region of A) the CON-IPAP spectrum and C) the ${}^1\text{H}$ - ${}^{15}\text{N}$ HSQC spectrum of 4Q (black peaks) and 25Q (red peaks) with an indication of the resonances that experiences the largest shifts upon increasing the length of the polyQ tract.

Analysis of the secondary structure propensity (SSP) of the polypeptide was performed to characterize the conformational properties of 4Q and 25Q from their heteronuclear backbone chemical shifts (232) (Fig. 4-6A). For 4Q, it identified one region of sequence with significant α -helical propensity, between residues 50 and 73 that contains the polyQ tract. For 25Q, the results were very similar for most of the sequence, with the exception of the polyQ region (residues 55-80) as elongation of the tract from four to 25 residues increased very substantially the α -helical propensity of this region, which reached values compatible with a fully-formed α -helix.

The backbone dynamics of 4Q (Fig. 4-7) and 25Q (Fig. 4-8) were studied by measuring ^{15}N relaxation data. ^{15}N longitudinal (R_1) and transverse (R_2) relaxation rates, as well as steady-state heteronuclear $^{15}\text{N}\{^{15}\text{H}\}$ NOEs, were determined for all non-proline backbone amide ^{15}N at 16.4 T (700 MHz) (Figs. 4-6B). As it was the case for the SSP analysis, the one of ^{15}N relaxation rates provided almost identical results for 4Q and 25Q, except for the polyQ tract and the regions of sequence adjacent to it (Fig. 4-6B and Fig. 4-8). We found the R_2 values to exhibit a substantial dispersion and detected three regions of sequence characterized by faster than average transverse ^{15}N relaxation: those centered around the motifs $^{22}\text{GAFQNLF}^{28}$ and $^{55}\text{LLLLQQQQ}^{62}$ and $^{87}\text{TGYLVLD}^{93}$ in 4Q, the last corresponding to residues 108-114 in 25Q.

The $^{22}\text{GAFQNLF}^{28}$ motif corresponds to a region of sequence with low disorder propensity that in full-length androgen bound AR can interact with activation function two (AF-2) in the LBD (107, 109, 233). The second motif was identified by the SSP analysis as partially α -helical and it is therefore reasonable to attribute the high values of transverse ^{15}N relaxation that it displays to the relatively long local correlation times caused by helix formation. The third motif corresponds to a region of relatively high hydrophobicity with some propensity to form helical conformations. A comparison of the transverse ^{15}N relaxation rates of 4Q and 25Q reveals substantially higher values for the $^{55}\text{LLLLQQQQ}^{62}$ motif of the latter, which it is reasonable to attribute to the increase in secondary structure propensity arising from the elongation of the polyQ tract.

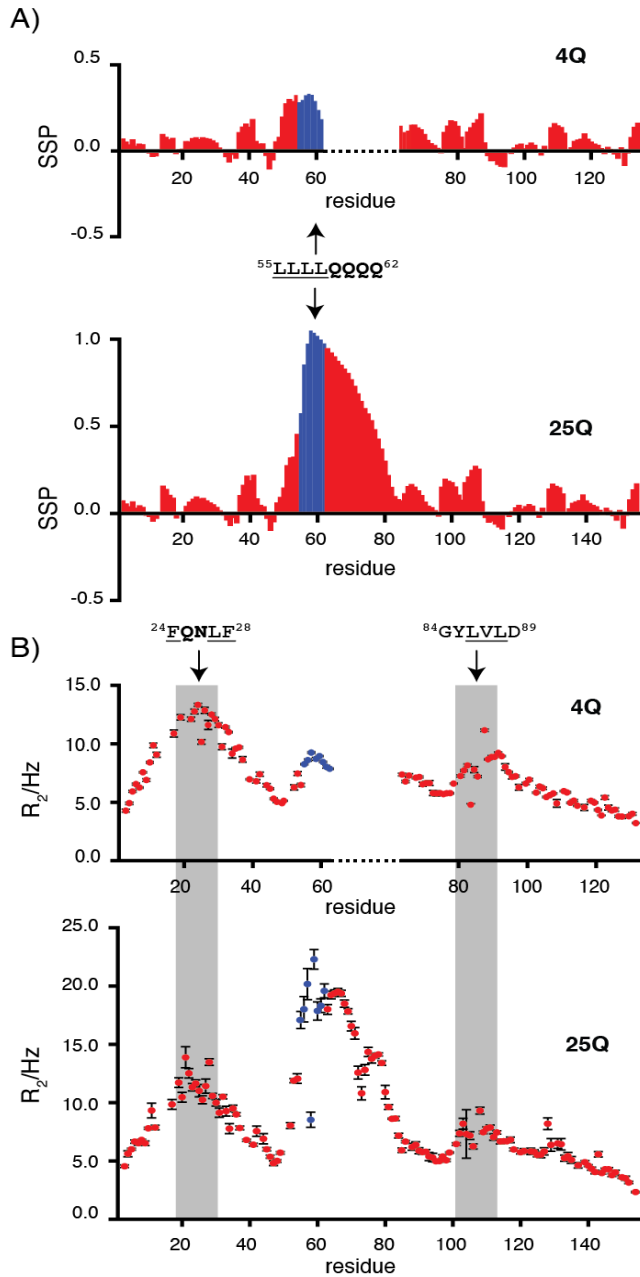


Figure 4-6: A) Secondary structure propensity of 4Q and 25Q based on an analysis of the C' , C^α , C^β and N backbone chemical shifts carried out by using the software SSP developed by Forman-Kay and co-workers (234). B) ^{15}N R_2 relaxation rates reported as a function of residue number for 4Q and 25Q. In A) and B) 4Q is reported with a gap of 21 residues, corresponding to the difference in length of 4Q and 25Q polyQ tract in order to facilitate their comparison; residues 55 to 62 of both 4Q and 25Q are shown in blue to highlight the influence of the length of the polyQ tract on their structural properties. In the motifs mentioned in the figure, hydrophobic residues are underlined and Q and N residues are shown in bold.

The details of backbone dynamics of 4Q are illustrated in figure 4-7. Three regions with higher R_2 values are indicated in colors. These are the similar regions that have already been discussed. As it is indicated, heteronuclear NOE and R_1 values are not as sensitive as R_2 values, as previously mentioned in Materials and Methods. The NOE values in this range are a clear indication of an IDP, which in this example comprises small regions with more structural propensities and higher rigidity. Increasing the number of Q, exclusively affect the $^{55}\text{LLLLLQQQ}^{62}$ region, with no effect on rest of construct.

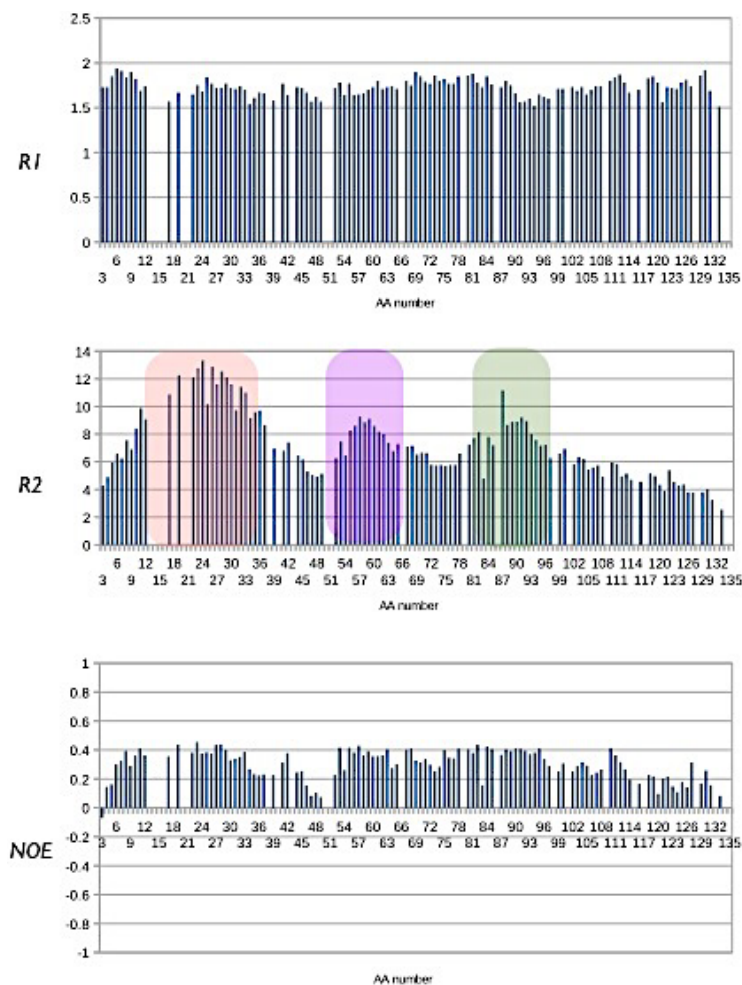


Figure 4-7: ^{15}N R_1 , R_2 relaxation rates and hetero-nuclear NOE reported as a function of residue number for 4Q. Among these data only R_2 is sensitive for our study. Three regions with higher R_2 have been indicated in pink ($^{24}\text{FQNLF}^{28}$), violet (polyQ) and green $^{87}\text{TGYLVLD}^{93}$. The higher R_2 is an indication of the presence of more rigid and structured regions in these motifs.

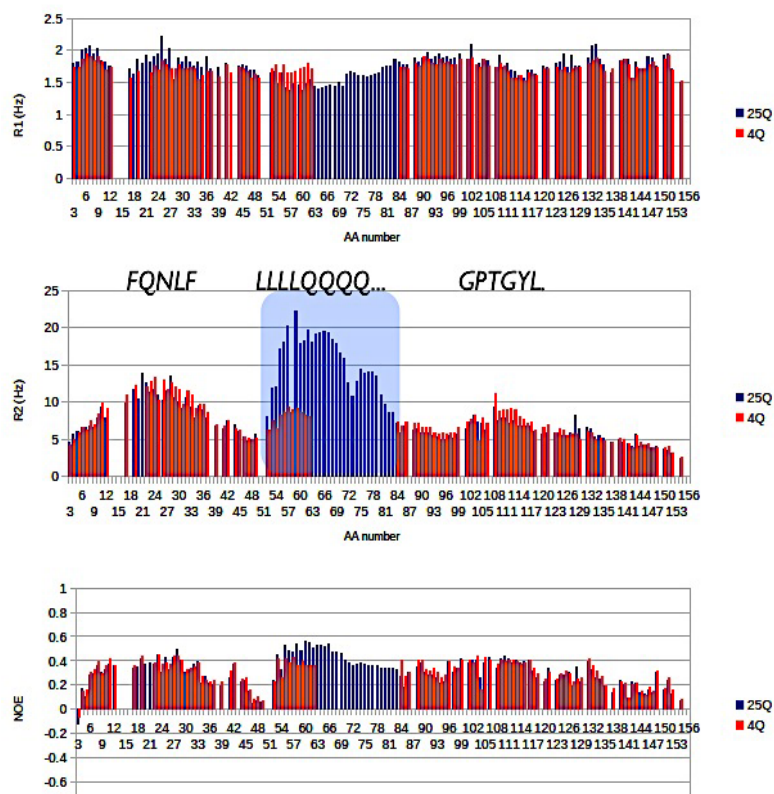


Figure 4-8: ^{15}N R_1 , R_2 relaxation rates and hetero-nuclear NOE of 4Q in red and 25Q in blue reported as a function of residue number for 4Q and 25Q. Among these data only R_2 is sensitive for our study. In a comparison between 4Q and 25Q, the elongation in polyQ region shows the only difference in R_2 rate, while the rest of the construct looks completely identical. The higher R_2 in $^{55}\text{LLLQQQQ}^{62}$ in 25Q compared to 4Q is another validation of the presence of an α -helical structured region in this motif in 25Q.

4.2.4 Identification of the residues undergoing structural changes during oligomerization of 25Q

The remarkable similarity of the ^1H - ^{15}N HSQC spectra of SDP and UID samples of 25Q, allowed us to transfer the ^1H and ^{15}N backbone chemical shift assignments from the latter to the former and therefore to identify the residues of 25Q that underwent changes in chemical environment during oligomerization (Fig. 4-9). This analysis revealed that the regions of sequence where these residues are found are $^{22}\text{GAFQNLFQSVREVIQ}^{36}$, $^{139}\text{GAAVAAS}^{145}$ and, to a lesser extent, the polyQ tract itself. $^{22}\text{GAFQNLFQSVREVIQ}^{36}$, that contains the $^{22}\text{GAFQNLF}^{28}$ motif with fast ^{15}N transverse relaxation discussed in the previous section, has a high content in glutamines and hydrophobic residues, a high propensity to fold, a relatively high helical propensity and a high AABUF. $^{139}\text{GAAVAAS}^{145}$ is a region of

relatively high helical propensity and low AABUF and propensity to fold (Fig. 4-1B). The polyQ tract is a region of high helicity in Q25 and relatively high AABUF.

A detailed analysis of the temporal evolution of the position and intensity of the peaks of the HSQC reveals that the different regions of sequence undergo conformational changes of different nature that occur with different rates. The first of these changes, which leads to a drop of signal intensity ($I/I_0 \approx 0.5$) that is clearly noticeable after two days of incubation, affects the first of these motifs and, to a lesser extent, the C-terminal one ($I/I_0 \approx 0.65$). Incubation of the sample for two additional days leads to further decreases in the signal intensity of these motifs to values of I/I_0 of 0.30 and 0.45, respectively, and to the observation of a modest ($I/I_0 \approx 0.70$) decrease in the region defined by residues 50 to 85, which includes the polyQ tract and the residues flanking it. Further incubation for a total of seven days leads to important decreases in intensity of the signals corresponding to residues in the N-terminal motif, which in many cases can no longer be observed ($I/I_0 \approx 0$), and to an enhancement of the effects observed in the polyQ tract and the C-terminal motif ($I/I_0 \approx 0.6$ and 0.3) (Figs. 4-9B and 4-10A).

As already mentioned the resonances of a small number of residues split into two or in one case more peaks in addition to decreasing in intensity. These resonances, which are among those that decrease in intensity the earliest, correspond to the motif $^{29}\text{QSVREVIQ}^{36}$, which is rich in both Gln and hydrophobic residues and has a predicted high helical propensity (Fig. 4-1B). This indicates that oligomerization proceeds via the formation of species with well-defined chemical shifts stabilized by intermolecular interactions involving side chains of this region of sequence. A particularly striking case is Val 31, shown in detail in Figure 4-9C, which evolves via the formation of at least one intermediate with chemical shifts that have values between those of the monomer and those observed at the end of incubation period (Fig. 4-9).

A comparison of the ^1H - ^{15}N HSQC spectra and DLS results obtained for SDP and UID 25Q shortly after sample preparation suggests that the kinetically stable samples contain oligomeric species that are in intermediate exchange and, therefore, in apparent equilibrium with monomeric 25Q: although the cross-peaks in the ^1H - ^{15}N HSQC spectra are in the same positions, allowing the transfer of the assignments, the intensities of the resonances of the sample treated with urea are markedly lower than those of the spectra of samples that were disaggregated, indicating that oligomers are present and in intermediate exchange with monomeric 25Q. In addition, whereas the Z-average of the SDP samples was very low shortly after dissolution (*ca.* 8 nm), but quickly increased as oligomerization took place, that of the UID samples had a relatively low value of *ca.* 100 nm and hardly increased during incubation, indicating that the sample contained oligomers but that their concentration increased very slowly.

These observations strongly suggest that UID generates a kinetically stable distribution of oligomeric species that can be used to study the nature of inter-

molecular interactions that stabilize the oligomers. With this purpose, we acquired 2D ^1H - ^{15}N HSQC spectra of UID samples at various concentrations (125, 250 and 500 mM) and observed that the relative intensity of the resonances of residues belonging to the motif $^{22}\text{GAFQNLFQ}^{29}$ decreased with concentration (Fig. 4-11A). Further measurements of ^{15}N transverse relaxation rates (R_2) at equivalent 25Q concentrations confirmed that the decrease in signal intensity was indeed due to the concentration-dependence of the ^{15}N transverse relaxation rates, which we attribute to the contribution of exchange processes (R_{ex}) due to oligomerization (Fig. 4-11B).

In summary our results indicate that all regions of sequence that are involved in the early stages of AR oligomerization in SBMA have predicted helical propensity and, in addition, that monomeric Q25 is in equilibrium with oligomeric species stabilized by inter-molecular interactions established by residues in the motif $^{22}\text{GAFQNLFQ}^{29}$.

We have characterized the residues that experience chemical shift perturbations during the oligomerization of 25Q in a SDP sample (Figs. 4-9 and 4-10A). The region of sequence where most of such residues are found is the motif $^{30}\text{SVREVIQ}^{36}$, of predicted high helicity, that is part of the region $^{22}\text{GAFQNLFQSVREVIQ}^{36}$ and is directly adjacent to the motif $^{22}\text{GAFQNLFQ}^{29}$. That the species with different chemical shifts in the $^{30}\text{SVREVIQ}^{36}$ motif are in slow exchange with monomeric 25Q indicates that they correspond to well-defined oligomeric species of a size that is compatible with their characterization by solution NMR. We conclude that the second step in the oligomerization of 25Q is the formation of oligomers stabilized by inter-molecular interactions involving mainly residues in the motif $^{30}\text{SVREVIQ}^{36}$ (Fig. 4-10B).

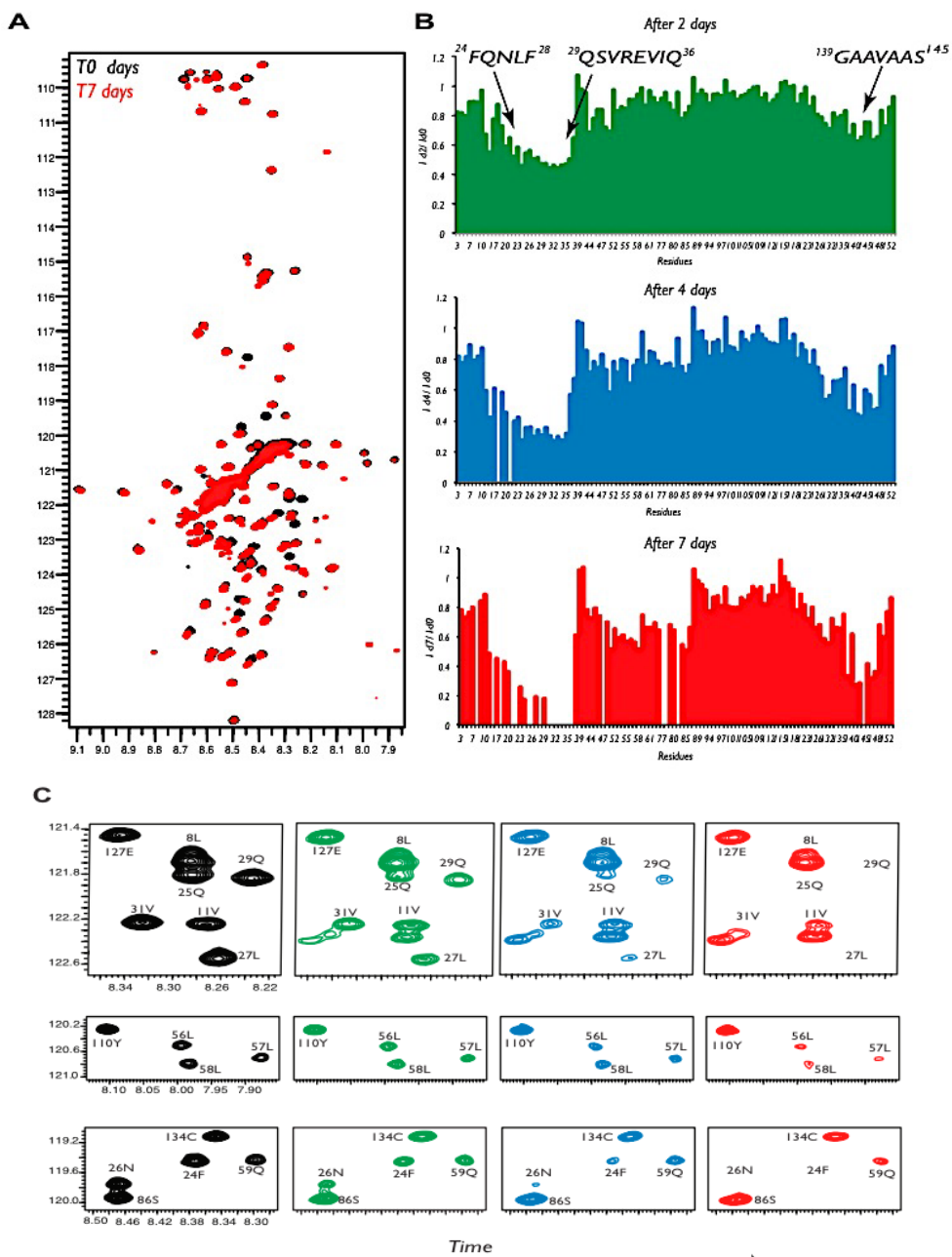


Figure 4-9: A) Overlay of the ^1H - ^{15}N HSQC spectra of a freshly prepared $50\mu\text{M}$ ^{15}N -labeled 25Q sample (black) and after seven days incubation at 37°C (red). B) Intensity ratios (I/I_0), where I is the normalized intensity of 25Q resonances after incubation at 37°C for two (green), four (blue) and seven (red) days and I_0 is the normalized intensity of 25Q resonances at time zero. The amino acid sequence of the most affected regions is indicated. C) Expansions of the ^1H - ^{15}N HSQC spectra of ^{15}N -labeled 25Q ($50\mu\text{M}$) at time zero (black) and after incubation at 37°C for two (green), four (blue) and seven (red) days highlighting the behavior of selected 25Q resonances.

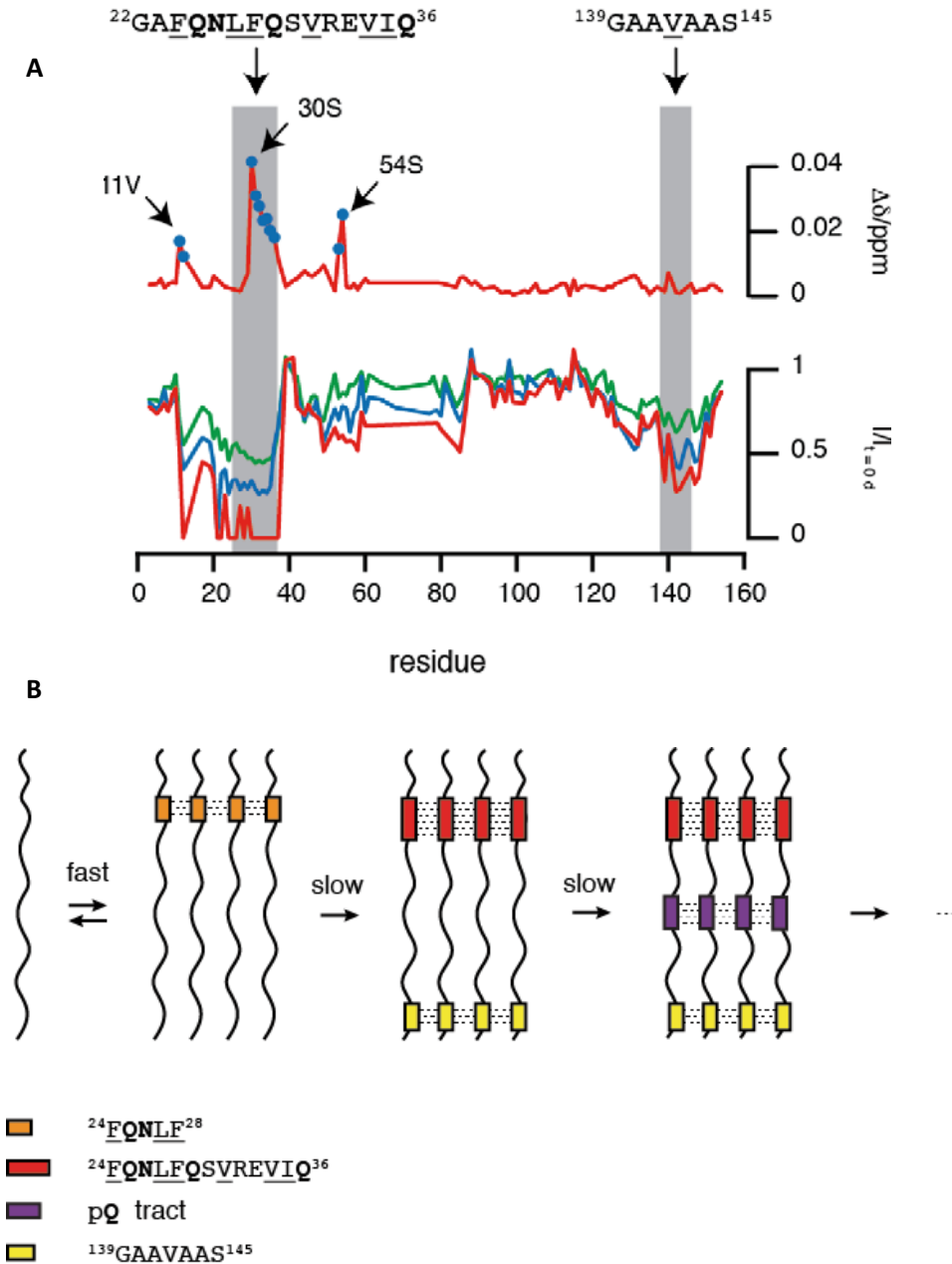


Figure 4-10: A) Perturbation of the chemical shifts ($\Delta\delta$, top) observed after 7 days of incubation and decreases in signal intensity, relative to that after sample preparation ($I/I_{t=0d}$, bottom) reported as a function of residue number and as a function of time. $\Delta\delta$ values shown as blue dots correspond to the difference in chemical shift between two peaks in slow exchange originating from a peak of monomeric 25Q that split during oligomerization. B) Scheme of the mechanism of oligomerization proposed for the N-terminal AR fragment based on the results obtained in this work.

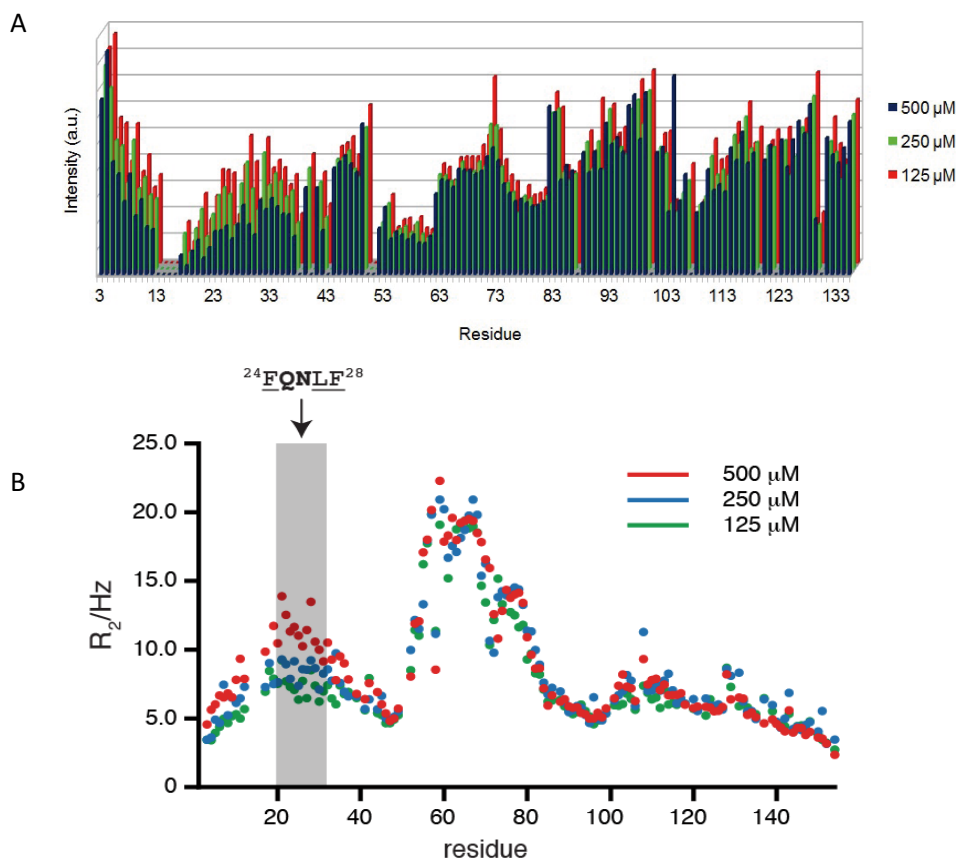


Figure 4-11: A) Peak signal intensities and B) ^{15}N R_2 relaxation rates for 25Q reported as a function of residue number at three different protein concentrations: 125, 250 and 500 μM . All the regions of construct show similar behavior in different concentrations, except for $^{24}\text{FQNLFQSVREVIQ}^{36}$ region that shows decrease in signal intensity and increase in R_2 . This is an indication of intermolecular interactions between monomers of 25Q. The error bars have been removed from the figure to facilitate the comparison and are included as supplementary information.

4.2.5 Predictions of the structural properties of 4Q and 25Q by bioinformatics tools

A thorough characterization of 25Q and 4Q was previously explained in detail. In the following, we add some supplementary information, which helped us with a better understanding of properties of 25Q and 4Q.

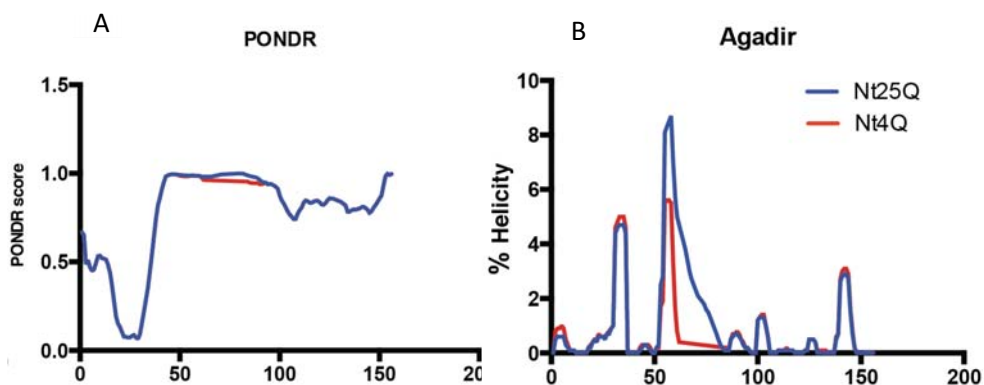


Figure 4-12: Disorder (PONDR) and % of helicity (Agadir) prediction of 25Q and 4Q have been indicated. (A) The PONDR score of $^{24}\text{FQNLF}^{28}$ is well below the 0.5 median normally associated with a tendency toward being structured, the PONDR score of rest of the falls above the 0.5 median, suggesting a negligible tendency toward order, whereas in the (B) Agadir predictor indicated two regions with higher helical propensity, $^{24}\text{FQNLF}^{28}$ and $^{55}\text{LLLLQQQ}^{62}$, both in 4Q and 25Q.

The interesting observation of predictors' data on disorder properties (Fig. 4-12A) and helicity (Fig. 4-12B) did not agree with our experimental data. The PONDR predictor indicates the disorder properties of a construct (235). In both 4Q and 25Q, the PONDR score of only $^{24}\text{FQNLF}^{28}$ is below the 0.5 median that is normally associated with a tendency toward being structured, in contrast to Agadir, which indicated in both 4Q and 25Q two regions with higher helical propensity. Thus, this it is to note that our experimental data does not support the predictors and this could be an example of flaws in the predicted data in structural biology.

4.2.6 4Q construct temperature dependence behavior observed by NMR

A series of 2D ^1H - ^{15}N HSQC and CON-IPAP NMR spectra were acquired on the 4Q construct at different temperatures, in the range 278-308 K, with steps of five degrees. The spectra are reported in Figures 4-13 and 4-14.

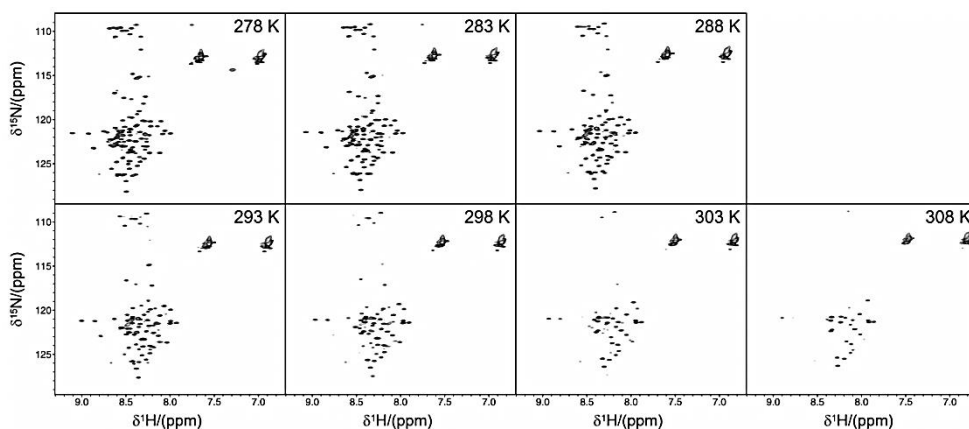


Figure 4-13: 2D ^1H - ^{15}N HSQC NMR spectra acquired at different temperatures (from 278 to 308 K).

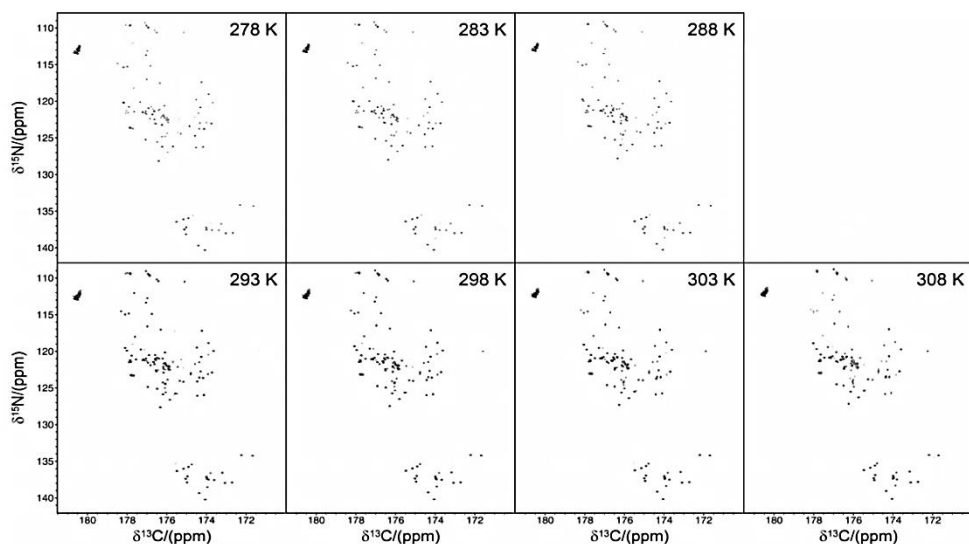


Figure 4-14: 2D CON-IPAP NMR spectra acquired at different temperatures (from 278 to 308 K).

As clearly visible, the quality of the 2D ^1H - ^{15}N HSQC spectrum is reduced with increasing temperature, due to the faster chemical exchange rate of amide protons with the solvent ones. Instead, the CON-IPAP NMR spectrum, insensitive to exchange processes benefits from higher temperature for the improved molecular tumbling.

4.3. Summary

Our data summarizes itself into two main findings. By using the nuclear magnetic resonance approach, two constructs of N-terminal domain of AR with 4Q and with 25Q were successfully assigned. Backbone amide assignment and information of chemical shifts of different nuclei have illustrated that although the whole polyQ diseases field believes in polyQ adapting a random coil structure, our data showed for the first time that in 4Q, the ⁵⁵LLLL⁵⁸ region just adjacent to polyQ has some helical propensity, while when the number of the Q was increased, the whole region -not only the polyQ tract but also the ⁵⁵LLLL⁵⁸ region- changes its α -helical propensity extensively from 0.4 to almost 1. Therefore, we have a clear evidence that N-terminal domain of AR comprises a polyQ part (even in Wt tract size), which forms a helix.

In context of full-length AR, the presence and formation of this helix is critical for AR intrinsic function. AR as a transcription factor is dependent on many protein-protein interactions. We suggest, as other studies have also shown, that the helix is involved in coiled-coil interactions with other proteins capable of coiled-coil formation. Thus, we assume, that in case of higher number of Q, the structural properties of this helix might change and could directly affect its binding partners. The new coiled-coil partners replace the older ones and this could explain the cell toxicity of CAG repeats due to a higher number of Q.

The second result, but as interesting as the first one, was the sequence determination of AR-NTD that initiates the oligomerization of AR and consequently brings the protein to aggregated state. We identified a motif in a distant N-terminal flanking region to polyQ. This region does not show any secondary structure propensities similar to what we have illustrated in polyQ region, whereas it has specific sequence and structural properties, which make it highly prone to form intermolecular interactions. These regions of interest, ²⁴FQNLF²⁸ and its proximate flanking regions are the main regions in 25Q and 4Q that due to their hydrophobic nature and their helical propensities initiate the oligomerization. The first hints for intermolecular interactions originated from signal intensity loss in this region in the time evolution studying of aggregation. Relaxation experiment data introduced this region as the least flexible region in the whole construct, suggesting the presence of intermolecular interactions. And finally, the data on concentration dependent signal intensity loss and respective relaxation rates comparison validated the hypothesis that the ²⁴FQNLF²⁸ region initiates the intermolecular interaction and any interference with this interaction could be a potential strategy for drug discovery.

The background of the page is a solid dark gray color. Overlaid on this background is a pattern of numerous overlapping circles of varying shades of gray, creating a textured, cellular appearance. The circles are arranged in a somewhat irregular grid, with some overlapping significantly more than others.

Chapter 5. Interaction of Hsp40 and Hsp72 with AR

5.1. Introduction

5.2. Results

5.3. Summary

Chapter 5. Interaction of Hsp40 and Hsp72 with AR

5.1. Introduction

A network of heat shock proteins (Hsps) mediates cellular protein homeostasis and plays a fundamental role in preventing aggregation associated with neurodegenerative diseases. An orchestrated action of conserved families of molecular chaperones is necessary to keep the balance between protein folding and misfolding and to provide the cell with a protein homeostasis machinery.

Spinal and bulbar muscular atrophy (SBMA) is an inherited motor neuron disease caused by the expansion of the polyQ tract within the AR. The nuclear inclusions consisting of the mutant AR protein are characteristic and combine with many components of ubiquitin–proteasome and molecular chaperone pathways (236), raising the possibility that misfolding and altered degradation of mutant AR may be involved in the pathogenesis. It has been reported that the overexpression of Hsp chaperones reduces mutant AR aggregation and cell death in a neuronal cell model (176). To determine whether increasing the expression level of chaperone improves the disease manifestations in a mouse model, a group in the University of Nagoya crossbred SBMA transgenic mice with mice overexpressing the inducible form of human Hsp70 (237). They demonstrated that high expression of Hsp70 markedly ameliorated the motor function of the SBMA model mice. In double-transgenic mice, the nuclear-localized mutant AR protein, particularly that of the large aggregates with positive IC2-AR staining, was significantly reduced. Monomeric mutant AR was also reduced in amount by Hsp70 overexpression, suggesting the enhanced degradation of mutant monomeric AR with elongated polyQ (39). These findings suggest that Hsp70 overexpression ameliorates SBMA phenotypes in mice by reducing nuclear-localized mutant AR, probably caused by enhanced mutant AR degradation (163). This study provides the basis for the development of an Hsp70-related therapy for SBMA and other polyQ diseases.

In SBMA, nuclear inclusions (NIs) containing mutant AR have been observed in the brainstem motor-nuclei, spinal motor neurons and some other organs (17). Such neuronal inclusions are common pathological features in polyQ diseases and are also co-localized with many components of the ubiquitin–proteasome and molecular chaperones (23, 75, 163, 238), raising the possibility that misfolding and altered degradation of the mutant protein is involved in the pathogenesis of SBMA as well as other polyQ diseases (239, 240). Furthermore, these chaperones and proteasome would facilitate refolding or proteolysis of the mutant protein and may play a role in protecting neuronal cells against the toxic properties of the expanded polyQ (176, 241).

Hsp70 overexpression has been reported to enhance the solubility and degradation of mutant AR (162). Hsps have also been shown to suppress aggregate formation and cellular toxicity in a wide range of polyQ disease models (238, 241). Recently, overexpression of the inducible form of rat Hsp70 ameliorated neurologic deficits and the neuronal degeneration of spinocerebellar ataxia type 1 (SCA1) transgenic mice, whereas the NIs consisting of mutant ataxin-1 were not reduced (237).

In addition to Hsp70, which plays a critical role in various neurodegenerative diseases including polyQ diseases, one of its closest functional partners, the Hsp40 family of molecular chaperones, (DnaJ) contributes to suppressing the toxicity of various amyloid proteins. The evolutionarily conserved J domain protein, DNAJ/Hsp40 functions primarily by stimulating the adenosinetriphosphatase (ATPase) activity of other chaperones, such as Hsp70. Higher eukaryotes have a variety of Hsp40 family members to regulate the substrate specificity of chaperones (160). A few types of Hsp40 proteins have been shown to dramatically suppress toxicity in polyQ disease models.

Overexpression of Hsp70 or combination of Hsp70 and Hsp40 reduced aggregate formation and provided cellular protection, which suggests that Hsp70 and Hsp40 might function together in chaperoning aggregation-prone proteins. The essential molecular chaperones are highly efficient in selectively recognizing misfolded proteins and maintaining them in soluble states. Hsp110 forms high-molecular weight complexes with Hsp70, and facilitates the nucleotide exchange of Hsp70. Recent studies have suggested that Hsp110 associated with protein aggregation and introducing Hsp110 prevented the toxicity of aggregation-prone proteins (242). As the Hsp70 homolog, it is possible that Hsp110 functions with Hsp40 in preventing the toxicity of aggregation-prone proteins. Recent studies have shown that 70 kDa heat-shock cognate protein cb (HSC70cb) is a family member of Hsp110, interacts physically and genetically with DNAJ-1 *in vivo* and that HSC70cb is necessary for DNAJ-1 to suppress polyQ-induced cell death in *Drosophila*. Co-expression of DNAJ-1 and HSC70cb had a dramatic protective effect on the polyQ disease and HD models and knocking down of hsc70cb, has largely reduced the protective effect of DNAJ-1 on polyQ toxicity. To date, all this data has provided us with an overview of the importance and role of molecular chaperones in reducing the toxicity of aggregates. There is also convincing data on co-functionality between Hsp70 and Hsp40 for their protein homeostasis activity. However, to date no demonstration of the mechanism of action of Hsps in a high-resolution exists. In this work, we have collected all the *in vivo*, in cell experiments data and analyzed the interaction of full length Hsp72 and DnaJB1 (Human Hsp40) with an AR N-terminal domain construct with a polyQ tract containing either 4 or 25 glutamine residues. We have studied these natural disease associated complexes with solution state NMR.

5.2. Experimental procedure

All [^1H , ^{15}N]-HSQC were acquired at pH 7.4 and 5°C on a 800 MHz spectrometer. Two ^{15}N -labeled AR constructs, 4Q and 25Q, were used. The 4Q and 25Q constructs comprise the first 135 and 155 amino acids of the N-terminal part of AR (Fig. 4.1). 4Q has only 4Q, whereas 25Q comprises the elongated polyQ tract with 25Q. We assigned both constructs (Figs. 4-4 and 4-5) and as well studied the interaction of Hsp40 and Hsp72 with AR. Jennifer Rauch in the laboratory of Prof. Gestwicki in UCSF expressed and purified Hsp70 and Hsp40.

5.3. Results

We identified the binding epitope of the molecular (co) chaperones Hsp40, Hsp72 in a 135-residue and 156-residue N-terminal constructs (residues 1 to 135 and 1 to 156, that we call 4Q and 25Q) (See Fig. 4-1), corresponding to an AR fragment, produced by proteolytic cleavage by Caspase 3, found in AR aggregates linked to SBMA (12, 13, 243).

Our strategy was to start with 4Q. Along the time, 4Q evolves less compared to 25Q (Fig. 4-3A) and therefore it would be easier to optimize the experimental conditions. The interaction studies were initially performed on 4Q and in following with 25Q. This set of experiment with 25Q was necessary, as recent NMR experiments carried out in our lab indicated that the secondary structure of the polyQ tract changes drastically with the number of Q residues (Fig. 4-6A).

This information was obtained by analyzing the [^1H , ^{15}N]-HSQC NMR spectrum of samples of 4Q and 25Q enriched in ^{15}N in the absence and in the presence of increasing concentrations of these potential interactors. More details of results are illustrated in the following.

5.3.1 Heat shock protein72 (Hsp72) recognizes the FQNLF motif and its flanking region in AR 4Q and 25Q

Previous studies showed that AR interacts with Hsp72 and the interaction is important for Wt AR function (244, 245), but it also plays a critical role in SBMA pathology (176). Therefore, we set out to map its binding site on AR by NMR. We acquired [$^1\text{H},^{15}\text{N}$]-HSQC spectra on ^{15}N -labeled AR-4Q and AR-25Q in absence and presence of different concentrations of unlabeled full-length Hsp72 provided by laboratory of Prof. Jason Gestwicki at the UCSF (Fig. 5-1). This method provides a fingerprint of all amine bonds of a protein, which allows mapping interaction sites by recording changes in position and/or intensity of individual peaks. A subset of signals decreased their intensities upon Hsp72 binding, including $^{24}\text{FQNLF}^{28}$ (Fig. 5-3). The spectra revealed an important decrease in intensity in region between residues 2-35 upon Hsp72 binding. We repeated the experiment with three different concentrations of Hsp72, while the concentration of 4Q and 25Q kept constant.

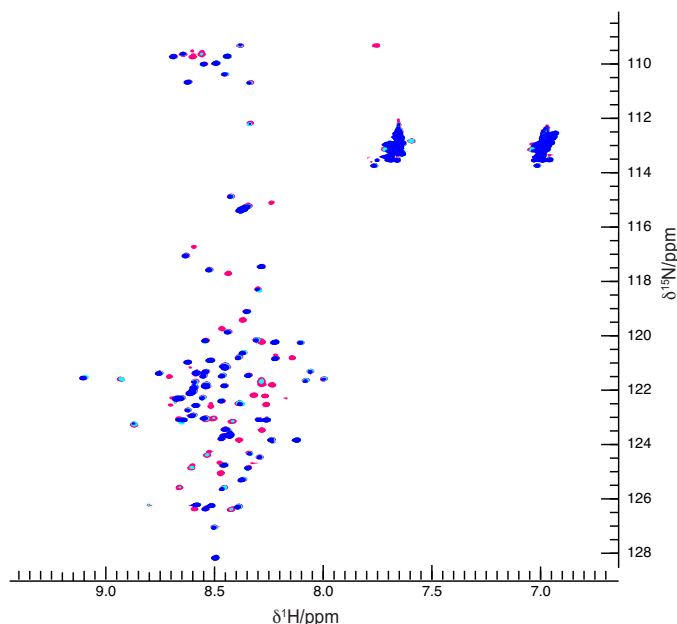


Figure 5-1: [$^1\text{H},^{15}\text{N}$]-HSQC of $33\mu\text{M}$ 4Q at 278 K in the absence (red) and in the presence of 1 (cyan) and 2.4 equivalents (blue) of Hsp72. Many resonances of 4Q are weakened in presence of Hsp72.

The reduction in signal intensity is a commonly observed phenomenon in protein NMR, which is typically caused by the combined effects of **(1)** the increase in molecular weight upon complex formation, which increases transverse relaxation

rates and thereby line width, and **(2)** the chemical exchange at the contact surface (246–249). The reduction in intensity is consistent with Hsp72-AR binding in the intermediate exchange time scale and also consistent with the affinity that was determined with fluorescence polarization substrate competition. Loss of peak intensity in this region confirms the direct interaction of AR with Hsp72 (Fig. 5-3 and 5-4).

The Hsp72 binding region contains the entire well-studied C-terminal binding site, including the ²⁴FQNLF²⁹. This motif is crucial for N/C interaction of AR upon binding of DHT (dihydrotestosterone) (105, 107, 108, 233) to its binding pocket in C-terminal domain (250, 251). This region shows most loss in signal intensity within the broad Hsp72 binding site (Fig. 5-4), indicating particular importance for the interaction in this region. This segment, which is representing the minimum in intensity, is characterized by large hydrophobic and aromatic side chains. It is already known from our NMR studies that this part of protein is highly prone to make intermolecular interactions (Fig. 4.11B). We therefore conclude that, it is not only the disordered nature of this motif that favors the binding of AR to chaperones, but also the presence of a hydrophobic pocket in this region make it highly favored for interaction with Hsps.

Interestingly though, our results do not validate the predictor's data (Fig. 5-2). We ran Hsp70 binding site prediction tool on the full length human AR sequence and it provided the following predicted Hsp70 binding sites (252).

Entry name	Position	Sequence	Score
sp P10275 ANDR_HUMAN	Androgen receptor	OS=Homo sapiens GN=AR PE=1 SV=2 53-59 SLLLLQ	16.6
sp P10275 ANDR_HUMAN	Androgen receptor	OS=Homo sapiens GN=AR PE=1 SV=2 54-60 LLLLQQ	14.7
sp P10275 ANDR_HUMAN	Androgen receptor	OS=Homo sapiens GN=AR PE=1 SV=2 103-109 TGYLVLD	11.4
sp P10275 ANDR_HUMAN	Androgen receptor	OS=Homo sapiens GN=AR PE=1 SV=2 433-439 WHTLF	18.8
sp P10275 ANDR_HUMAN	Androgen receptor	OS=Homo sapiens GN=AR PE=1 SV=2 550-556 DYYFPP	11.3
sp P10275 ANDR_HUMAN	Androgen receptor	OS=Homo sapiens GN=AR PE=1 SV=2 604-610 DKFRRK	11.4
sp P10275 ANDR_HUMAN	Androgen receptor	OS=Homo sapiens GN=AR PE=1 SV=2 713-719 VHVVKWA	13.9
sp P10275 ANDR_HUMAN	Androgen receptor	OS=Homo sapiens GN=AR PE=1 SV=2 742-748 MGLMVFA	12.2
sp P10275 ANDR_HUMAN	Androgen receptor	OS=Homo sapiens GN=AR PE=1 SV=2 746-752 VFAMGWR	19.7
sp P10275 ANDR_HUMAN	Androgen receptor	OS=Homo sapiens GN=AR PE=1 SV=2 796-802 WLQITP	19.8
sp P10275 ANDR_HUMAN	Androgen receptor	OS=Homo sapiens GN=AR PE=1 SV=2 807-813 MKALLLF	13.4
sp P10275 ANDR_HUMAN	Androgen receptor	OS=Homo sapiens GN=AR PE=1 SV=2 809-815 ALLLFSI	12.1
sp P10275 ANDR_HUMAN	Androgen receptor	OS=Homo sapiens GN=AR PE=1 SV=2 829-835 ELRMNYI	12.9
sp P10275 ANDR_HUMAN	Androgen receptor	OS=Homo sapiens GN=AR PE=1 SV=2 879-885 DLLIKSH	11.8

Figure 5-2: Prediction of DnaK (Hsp70 homologue) and full length human AR binding via molecular modeling and experimental data. The first three specified residues of interest in our construct are indicated in pink (Online software used from Limbo webpage (252)).

The data in Figure 5-2 shows that the prediction algorithms are not entirely reliable, as they didn't predict FQNLF region as the main binding motif of Hsp72 on AR. However, interestingly enough, two other interacting motifs on AR, though with much less affinity compare to FQNLF region, were predicted. These regions are indicated in red (SLLLLQQ and TGYLVLD). It is to note that the predictor was trained with the *E.coli* Hsp and it may be the reason that the specificity is different for human Hsp.

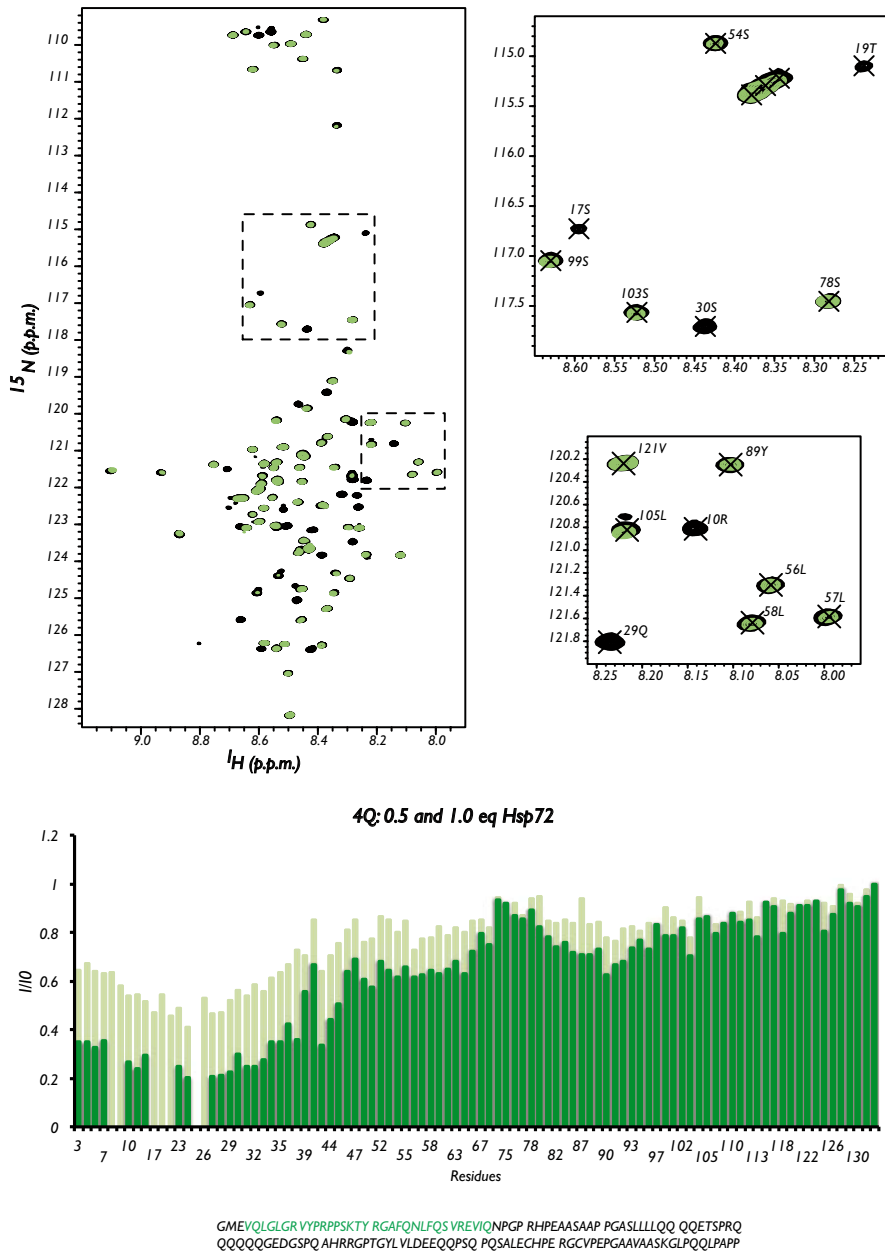


Figure 5-3: The top left panel illustrates that addition of Hsp72 decreases the intensity of the signals of the NMR spectrum of 4Q. The smaller panels on the right correspond to two regions of the spectrum, shown as boxes, which we wish to highlight. In these boxes, we show the assignment of the signals to illustrate how they decrease in intensity as the interactors are added. The plot at the bottom represents the intensities of the signals, relative to that of the free 4Q construct, in the presence of 0.5 and 1 molar equivalents of Hsp72.

GMEVQLGLGRVYPRPPSKTYRGA**FQNL**F**QSVREVIQ**N**PGPRHPEAASAAPPGAS**LLLL**Q**QQ****
QQQQQQQQQQQQQQQQQQQQQQ**Q**ETSPRQQQQQQG**EDGSPQA**HRRGPTGYLVLDEE
 QQPSQPQSALECHPERGCVPEPGA**AAVAASKGLP**QQLPAPP

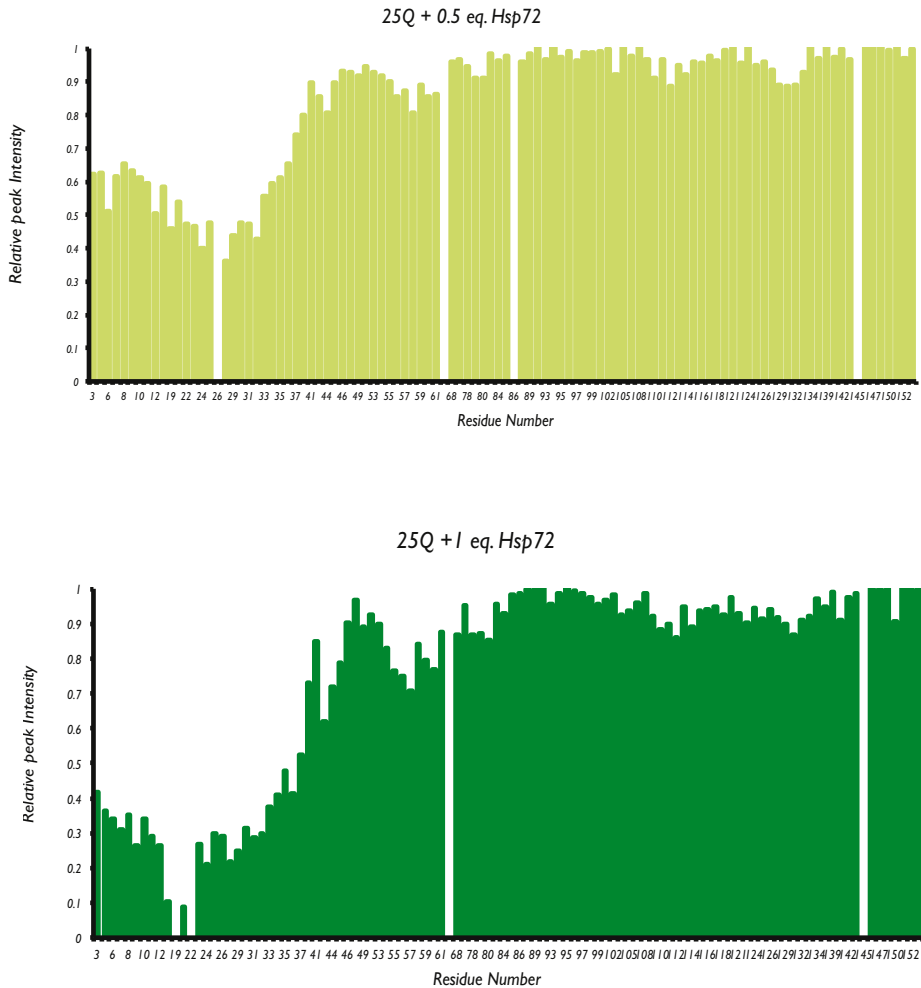


Figure 5-4: Ratio (I/I_0) of Normalized 25Q resonance intensities in the presence of 0.5 (top), 1.0 (bottom) molar equivalents of Hsp72. Hsp72 binds to $^{24}\text{FQNLF}^{28}$ with high affinity.

5.3.2 Heat shock protein 40 (Hsp40) binds less specifically to AR

As explained in the introduction, one of close partners of Hsp70 for its activity is Hsp40. There is evidence of the importance of the presence of both Hsp and their functions in different neurodegenerative diseases like Alzheimer’s disease as well as in polyQ diseases (164, 175) . Therefore, we used the similar approach of using NMR to understand the binding regions of Hsp40 to AR-4Q and AR-25Q. We followed the interaction of Hsp40 by acquiring $[^1\text{H},^{15}\text{N}]$ -HSQC spectra of ^{15}N -labeled proteins in absence and presence of different concentrations of Hsp40 (Fig. 5-5). Using the fingerprint of all the amide bonds of AR, we mapped the

interaction site by recording changes in the intensities of individual peaks (Figs. 5-6 and 5-7). Interestingly, some of the peaks overlapped with Hsp72, but some new subsets of peaks signal intensities also decreased upon Hsp40 binding (Fig. 5-3 and 5-6).

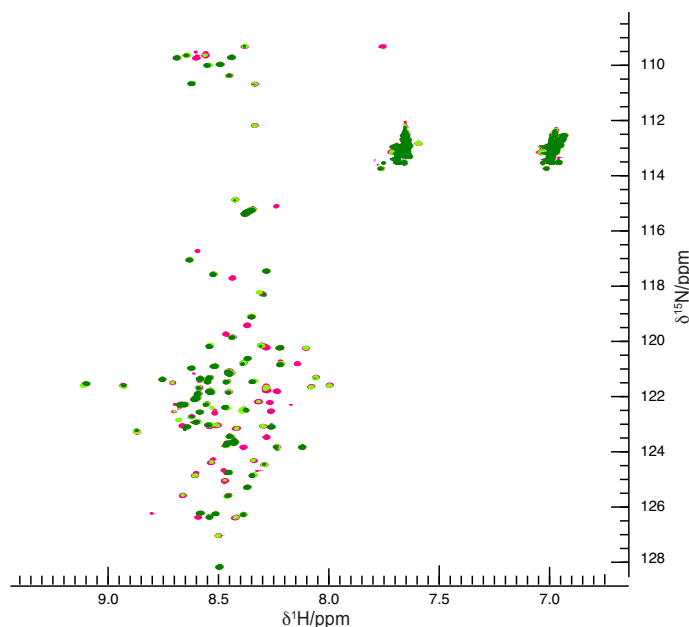


Figure 5-5: $[^1\text{H}, ^{15}\text{N}]$ -HSQC of $33\mu\text{M}$ 4Q at 278 K in the absence (red) and in the presence of 1 (light green) and 2.4 equivalents (dark green) of Hsp40. The effect of incubation with Hsp40 is similar to that caused by Hsp72 but stronger.

The binding regions of AR to Hsp40 include the motifs, $^{24}\text{FQNL}^{\text{F}28}$, $^{55}\text{LLLL}^{\text{L}58}$ and $^{106}\text{GPTGYLVL}^{\text{L}114}$. Upon binding of AR to Hsp40, the spectra revealed an important decrease in a concentration dependent manner in peak signal intensity in $^{55}\text{LLLL}^{\text{L}58}$ flanking region to polyQ tract and $^{106}\text{GPTGYLVL}^{\text{L}114}$ motif (Figs. 5-6 and 5-7). The decrease in intensity was higher in interaction with 4Q compared to 25Q (compare Fig. 5-6 and Fig. 5.7).

Interestingly, these two identified regions share some similarities; they both comprise hydrophobic residues with high AABUF values (Fig. 4-1B) and are shown to have similar relaxation properties (see Figs. 4-7 and 4-8). The reduction in peak signal intensity in these regions is also consistent with the Hsp40-AR binding affinity data determined by ITC (Fig. 5-13). Our data suggests that Hsp40 compared to Hsp72 shows less specificity for binding to its client protein. Hsp40 binds mainly to disordered regions and motifs with high content of hydrophobic residues.

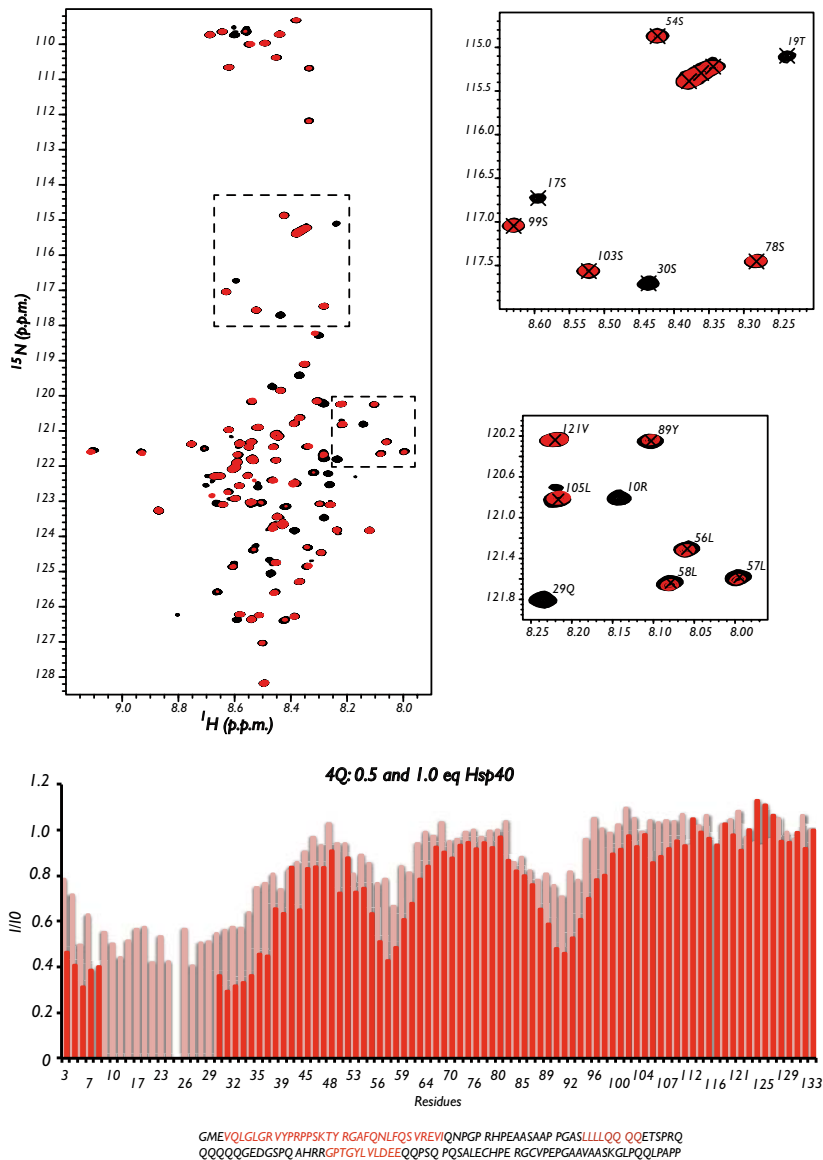


Figure 5-6: The top left panel illustrates that addition of Hsp40 decreases the intensity of the signals of the NMR spectrum of 4Q. The smaller panels on the right correspond to two regions of the spectrum, shown as boxes, which we wish to highlight. In these we show the assignment of the signals to illustrate how they decrease in intensity as the interactors are added. The plot at the bottom represents the intensities of the signals, relative to that of the free 4Q construct, in the presence of 0.5 and 1 molar equivalents of Hsp40. The peak intensity ratio shows that Hsp40 binds mainly in 3 regions, which includes the relatively hydrophobic motifs ²⁴FQNLF²⁸, ⁵⁵LLLL⁵⁸ and ⁸⁵GPTGYLVLD⁹³. The observed effects are similar to those observed for Hsp72 but stronger in ²⁴FQNLF²⁸ region.

GMEVQLGLGRVYPRPPSKTYRGAFQNLFFQSVREVIQNPGRHPEAASAAPPAS**LLLLQQQ**
QQQQQQQQQQQQQQQQQQQQQQEQETS**PR**QQQQQQG**EDG**SPQA**HRR****GPTGYLV**LDEE
 QQPSQPQSALECHPERGCVPEPGA**AAVA**ASKGLPQQLPAPP

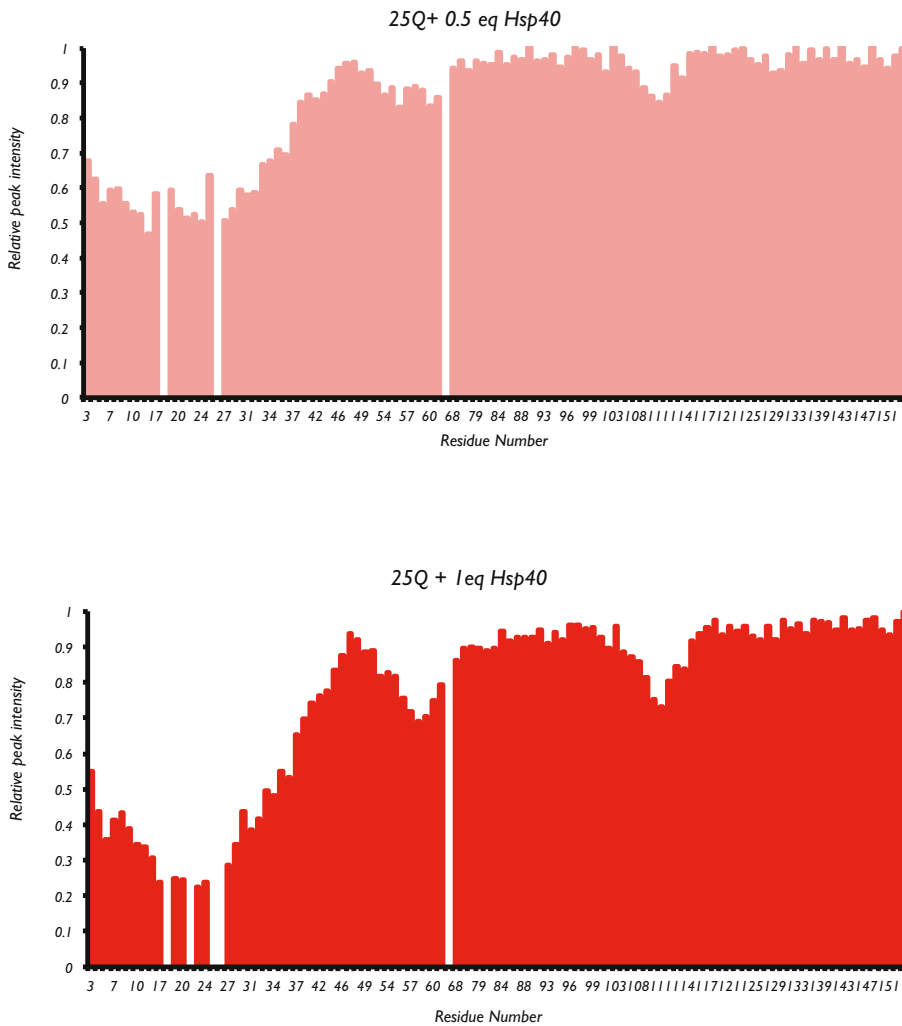


Figure 5-7: Ratio (I/I_0) of Normalized 25Q resonance intensities in the presence of 0.5 (top) and 1.0 (bottom) molar equivalents of Hsp40. The peak intensity ratio shows that Hsp40 binds mainly to ²⁴FQNLFF²⁸. Other hydrophobic motifs, such as ⁵⁵LLLL⁵⁸ and ¹¹¹LVL¹¹³ were also affected, however to a lesser extent to that of 4Q. The result indicates that the interaction of Hsp72 with the FQNLFF motif is not affected by an extension of the polyQ tract but the interaction of Hsp40 with the polyQ and the ¹⁰⁶GPTGYLVLD¹¹⁴ motif are instead affected.

Looking back at figure. 4-1B in chapter 4, it is interesting to note that these three identified regions have the highest AABUF value among all the residues. These observations suggest that hydrophobicity is the driving force for interaction of these chaperones with AR.

Previous studies have shown that Hsp40 and Hsp70 are partners and have complementary functions (175) (See Fig. 1.15) . Therefore, we studied the interaction of both Hsp72 and Hsp40 together with AR.

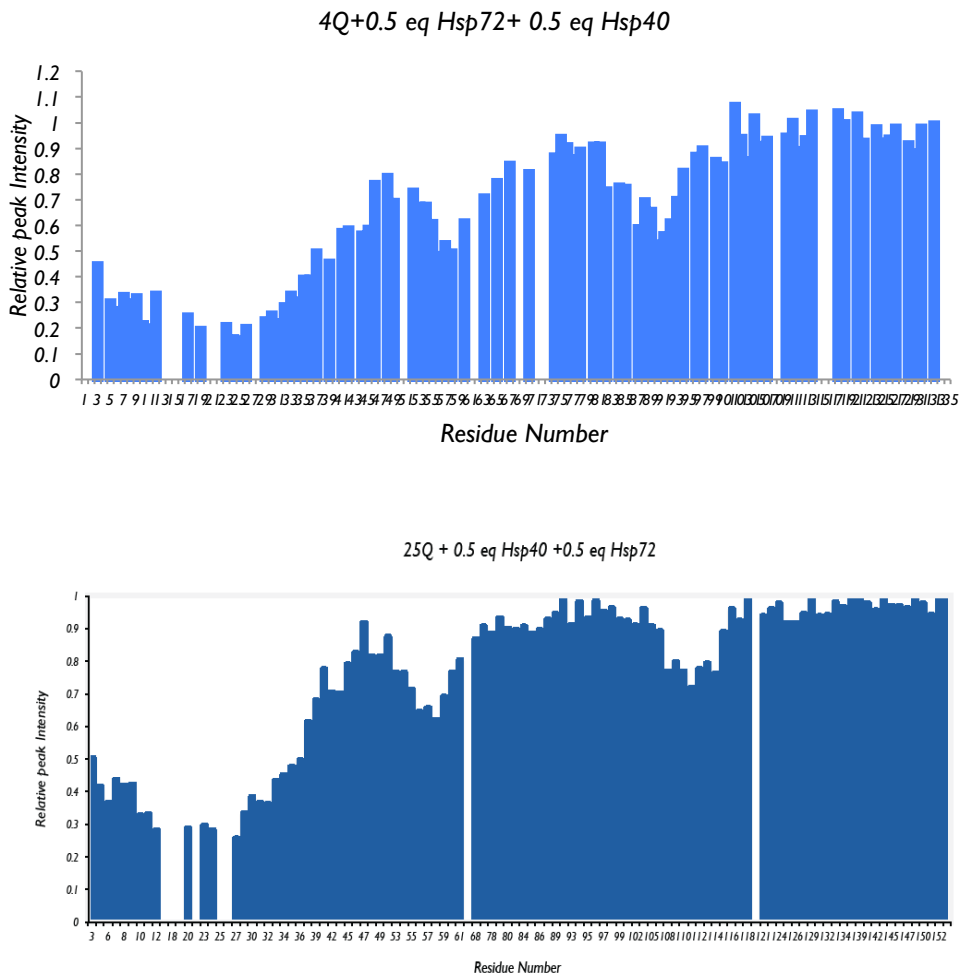


Figure 5-8: Ratios of normalized 4Q (top) and 25Q (bottom) cross-peaks intensities in the presence of 0.5 equivalents of both Hsp40 and Hsp72.

As illustrated in figure 5-8, Hsp40 and Hsp72 have additive binding affinity in both 4Q and 25Q (for 4Q, compare with Fig 5-6 and 5-3 in presence of 0.5 equivalent of Hsps and for 25Q, compare with Fig 5-7 and 5-4). This implies that in presence of both Hsp72 and Hsp40, the signal intensity loss is equal to addition of the conditions, when Hsp72 and Hsp40 were solely added. Therefore, we suggest that in this experimental conditions, there is no considerable increase or decrease in binding of any of these two Hsps, while the other binding partner is present. Our

data suggest that a clear cooperativity was not observed in this experimental conditions. To have a strong statement about the cooperativity of function of Hsp70 and Hsp40, more *in vitro* and in cell experiments will be necessary.

5.3.3 Presence of ATP does not significantly change the binding affinity of Hsp70

Hsp70 family of molecular chaperones has an essential role in folding and translocation of unfolded and folded peptides. While the general features of these activities are well studied, much less is understood about the regulation of these activities. Hsp70 interacts with non-native protein substrates in an ATP-dependent reaction cycle (253) regulated by J-domain proteins (an example of this is DnaJB1 or Hsp40) (254) and nucleotide exchange factors (NEFs) (170, 255, 256). Since the intrinsic ATPase activity of Hsp70 is very weak for stoichiometric coupling to the substrate binding-release cycle, J-proteins act as accessory factors for Hsp70 (257).

Thus, it was interesting to investigate the effect of ATP on the binding affinity of Hsp72 to AR (both 25Q and 4Q), using ATPgS and ATP. The pattern of NMR signals in ATP state changed for few of the peaks, whereas the signal intensities was identical to the conditions where no ATP was present, given the fact that the binding affinity of the ATP state did not change as shown in figure 5-9. Thus, we can not conclude that presence of ATP can modulate the interaction of Hsp72 and AR.

As it was illustrated in figure 5-10, the possible effect of presence of Hsp40 on ATPase activity of Hsp77 and its binding affinity to AR was studied. In presence of both Hsp40 and Hsp72, presence of ATP didn't induce any binding affinity changes in 4Q. A general overall decrease in binding was found in interaction between 4Q and Hsp72/Hsp40 complex. Similar observations in 25Q confirmed that in our experimental condition, presence and absence of ATP, although it induces few changes in peak positions, it does not significantly affect the interaction between AR and Hsp72/Hsp40 complex.

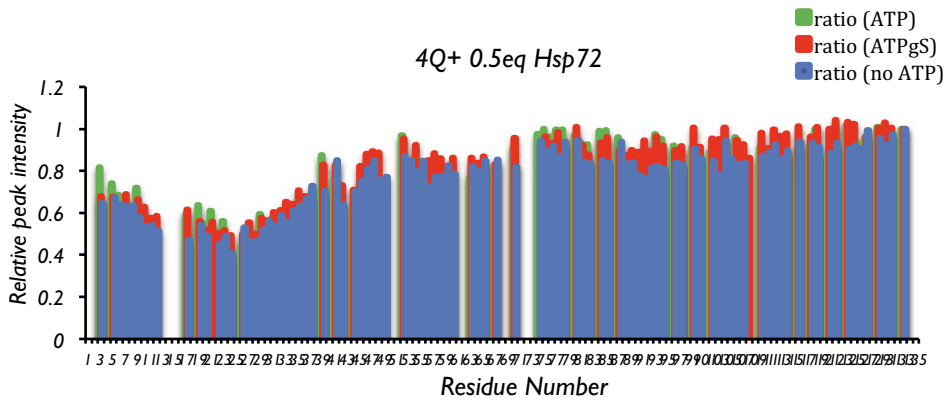


Figure 5-9: Normalized 4Q cross-peaks intensities in the presence of (top) both 0.5 equivalents of Hsp72 and Hsp40 and (bottom) 0.5 equivalents of Hsp72 in presence of ATP (1mM) and ATPgS (1mM).

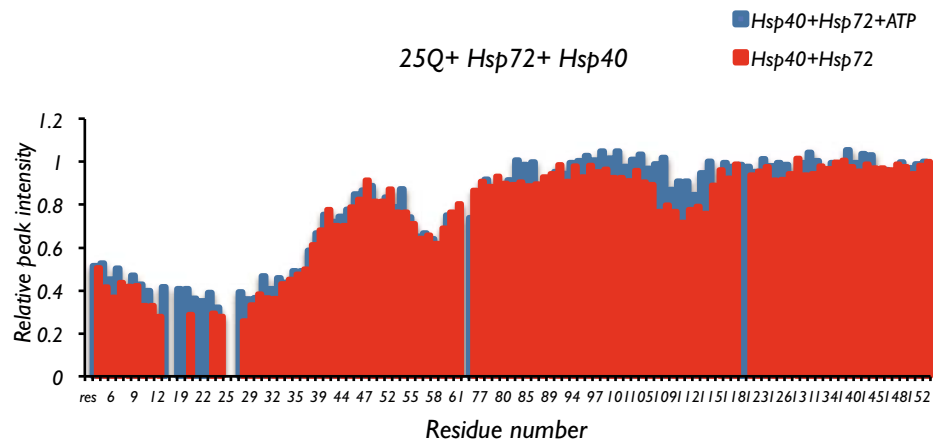
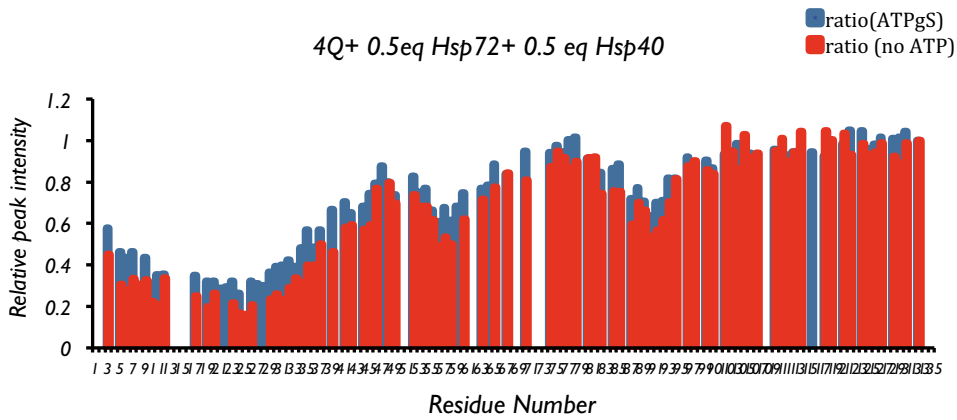


Figure 5-10: Normalized 4Q (top) and 25Q (bottom) cross-peaks intensities in the presence of both 0.5 equivalents of Hsp72 and Hsp40 in absence and presence of ATP (1mM).

Our findings support a model of chaperone machinery function, whereby Hsp40 and Hsp72 have collaborative roles in AR protein quality control (Fig. 5-11). Two main regions of interest, $^{24}\text{FQNLF}^{28}$ and $^{55}\text{LLLLQQQ}^{62}$, are main binding epitopes of Wt AR, as well as AR with elongated polyQ tract.

A monomeric AR comprises one LBD, one DBD and a large unfolded intrinsically NTD. In our model, based on our NMR data from chapter 4 and this chapter, we hypothesize that Hsp72 mainly binds to $^{24}\text{FQNLF}^{28}$ and its adjacent flanking region, while Hsp40 shares the same binding motif with but additionally binds to $^{55}\text{LLLLQQQ}^{62}$.

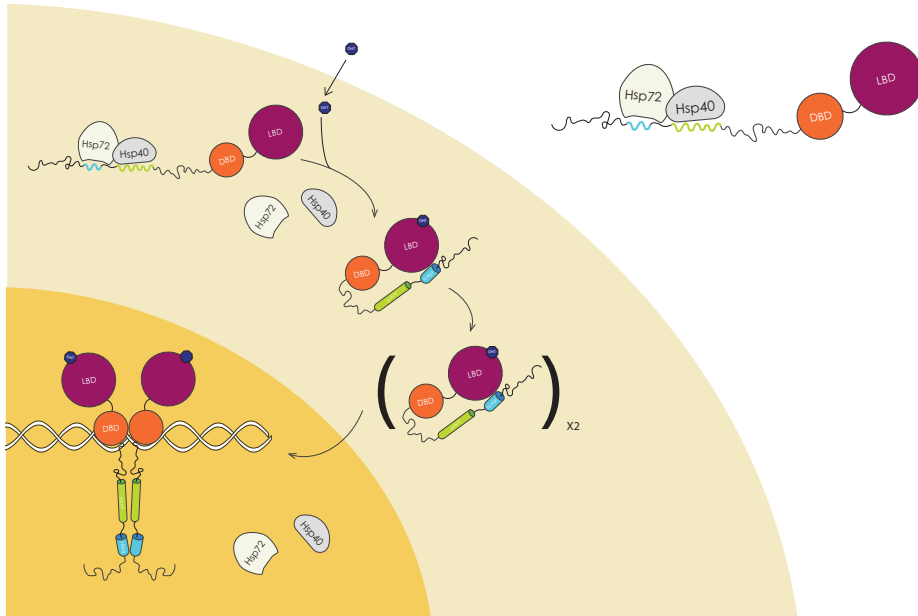


Figure 5-11: Our proposed model of interaction of the molecular chaperones Hsp40 and Hsp72 with the transactivation domain of AR A) Model for monomeric full-length AR with indicated LBD (ligand binding domain), DBD (DNA binding domain) and intrinsically N-terminal domain. B) Hsp Hsp72 and Hsp40 interact with AR intrinsically N-terminal domain in cytoplasm. Hsp72 and Hsp40 bind to NTD and to the respective binding motifs in blue ($^{23}\text{FQNLF}^{28}$) and in green ($^{54}\text{LLLLQQQ}^{61}$). In the presence of DHT and its binding to LBD, Hsps will be detached from AR and $^{24}\text{FQNLF}^{28}$ motif in NTD interacts with LBD. Consequently, 2 monomers of AR (either with heat to head or head to tail conformations) will move to nucleus. DBD will binds to DNA and DNA binding domains arrangements will induce the formation of head to head AR homo-dimers on DNA level.

Consistent with our model, we can explain that N/C interaction is necessary for the function of AR in presence of hormone in cytoplasm. This interaction plays a crucial role as the internal chaperone by binding to $^{24}\text{FQNLF}^{28}$ region and protecting it, when no Hsps present to protect the protein.

5.3.4 Biophysical characterization of binding affinity of Hsp40 and Hsp72 to 25Q and 4Q

We mapped the binding site of the full length human Hsp72 and Hsp40 on AR N-terminal domain. The complexes we have studied consisted of Hsp full-length proteins and a biologically relevant construct of AR truncated by Caspase cleavage, which is suggested to be the main species in SBMA aggregates (12, 13).

It was remarkable to show for the first time, how and where exactly Hsp40 and Hsp72 interacts with AR N-terminal domain. To illustrate in more details the binding affinity of these proteins to AR, Isothermal titration calorimetry (ITC) and fluorescence polarization substrate competition were used. Both techniques were well used to study protein-protein interaction (258–261).

Protein-protein interactions (PPIs) are emerging as novel therapeutic targets (262, 263), and there are ample examples of PPI disrupters based on antibodies, proteins or peptides. However, it is of interest to pursue “drug-like” small molecules as biochemical tools or potential compounds for therapeutic agents, to complement or replace the existing protein therapeutics. In our examples of protein-protein interaction, there are well-studied data on the importance of modification of Hsp interactions with AR, not only as therapeutic option in SBMA but also in prostate cancer (264).

5.3.5 Both Hsp72 and Hsp40 interacts with AR with $1\mu\text{M}$ affinity

In collaboration with Prof. Gestwicki laboratory in UCSF, Jennifer Rauch, a PhD candidate in his laboratory, performed ITC experiments with AR4Q and AR25Q with Hsp72 full length, SBD of Hsp72, BAG1, and DJA2 (Hsp40). AR was chosen as the cell reactant (20uM) and Hsps were concentrated to (200uM) for the syringe reactant.

The presence of several binding sites for both Hsp72 and Hsp40 is reflected in ITC experiments by a complex binding behavior, and therefore we refer to these binding affinities as “effective” affinity.

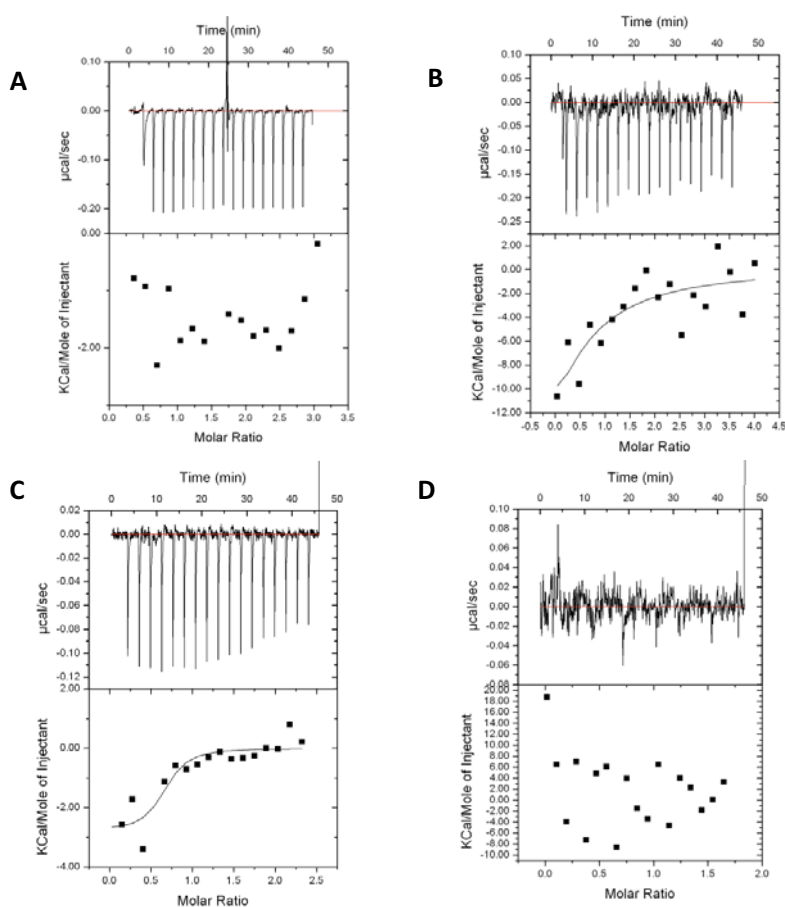


Figure 5-12: Hsp72 shows variable binding to AR4Q (A&B) and AR25Q(C&D). These low affinity interactions will need higher concentrations to determine accurate affinity values. We propose that the complex binding behavior of full-length Hsp72 and AR (multi binding motifs) can explain the low quality of ITC data and complicated curve fittings.

Varying results were seen for AR4Q-Hsp72 binding (Fig. 5-12). Due to the inherent weak nature of the interaction, apparent affinities ranged from no binding to 15 μ M (Fig. 5-12A) were observed. However, even binding was detected, overall signal was very low (Fig. 5-12B). Therefore, the affinity constants obtained should be taken with low confidence. While some heat is observed upon binding, higher concentrations of samples would be needed to fully elucidate binding affinities. Similar results for AR25Q-Hsp72 were achieved (Figure 5-12C & D). Higher concentrations of partners will be needed to report accurate binding affinities. Repeating the experiments with only substrate binding domain of Hsp72 indicated a more precise binding affinity information on interaction of both 4Q and 25Q with Hsp70 substrate binding domain (Fig. 5-15B). Moreover, this observation suggests that in the main binding motif of Hsp72 to AR, is the substrate binding motif.

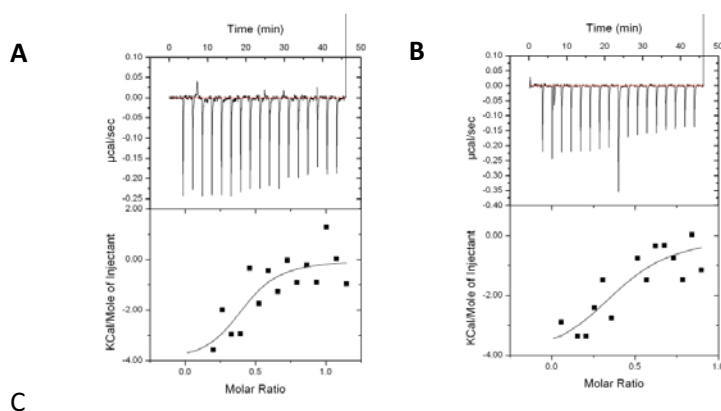


Figure 5-13: (A) DJA2-AR4Q ITC (B) BAG1-AR25Q ITC (C) Table of values from one replicate.

DJA2 (Hsp40) & BAG1 (another Hsp70 binding partner) binding to AR-4Q and 25Q showed more promising ITC results with affinities listed in Fig. 5-13. While signal was still low for these proteins as well, less noise in the control samples (syringe reactant into buffer) made data interpretation cleaner. (Fig. 5-13)

5.3.6 Fluorescence polarization substrate competition validates the interaction between Hsp72 substrate binding domain and AR

Due to complex binding behavior of both Hsp40 and Hsp72 and presence of multiple binding sites on AR, high error rates were observed in ITC studies. Therefore, FP assays were performed, assuming that AR binds to Hsp72 in the substrate-binding pocket, Jennifer Rauch was able to use a FP platform to look at competition between AR constructs and the known Hsp72 peptide binders.

The first experiment performed, used a tracer from a known Hsp72 client human leukocyte antigen (HLA). Hsp72 binds the HLA tracer with a $1 \pm 0.2 \mu\text{M}$. Keeping a constant concentration of Hsp72 and HLA tracer, AR4Q and AR25Q were titrated to compete for binding. Experiments were performed in duplicate and repeated twice on two independent days with an average IC_{50} listed below (Fig. 5-14).

IC_{50}

AR4Q: $2.6 \pm 0.5 \mu\text{M}$

AR25Q: $1.4 \pm 0.6 \mu\text{M}$

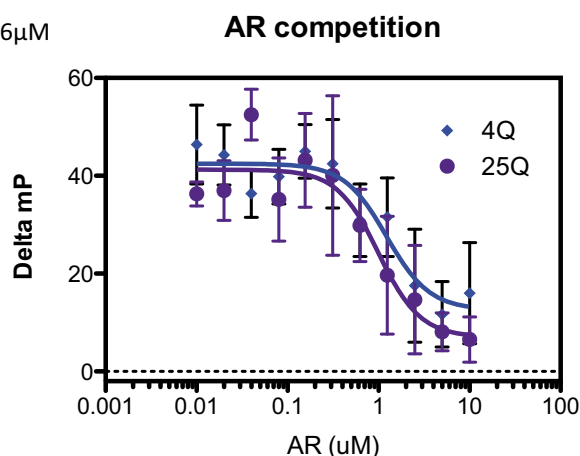


Figure 5-14: Fluorescence polarization substrate competition studies of interaction of AR with Hsp72. In independent experiments, 4Q and 25Q were titrated to compete for binding to Hsp72, while Hsp72 and HLA tracer were present in the solution.

In the second FP experiment, a tracer (LVEALY) from a known Hsp72 client Tau was used (265). This tracer binds with a large mP signal to the Hsp72 SBD (residues 394-540) and with an affinity of $1.5 \pm 0.3 \mu\text{M}$. AR25Q was highly potent in competing with this tracer, again supporting that AR25Q binds specifically to the SBD of Hsp72 in a canonical Hsp70-client protein interaction. These experiments were performed in duplicate and repeated on three independent

days with an average IC_{50} reported below (Fig. 5-15- upper panel). Indicating that SBD of Hsp72 interacts with AR, similar ITC experiments were performed with Hsc70 SBD (394-540) and more convincing results were obtained (Fig. 5-15 lower panel).

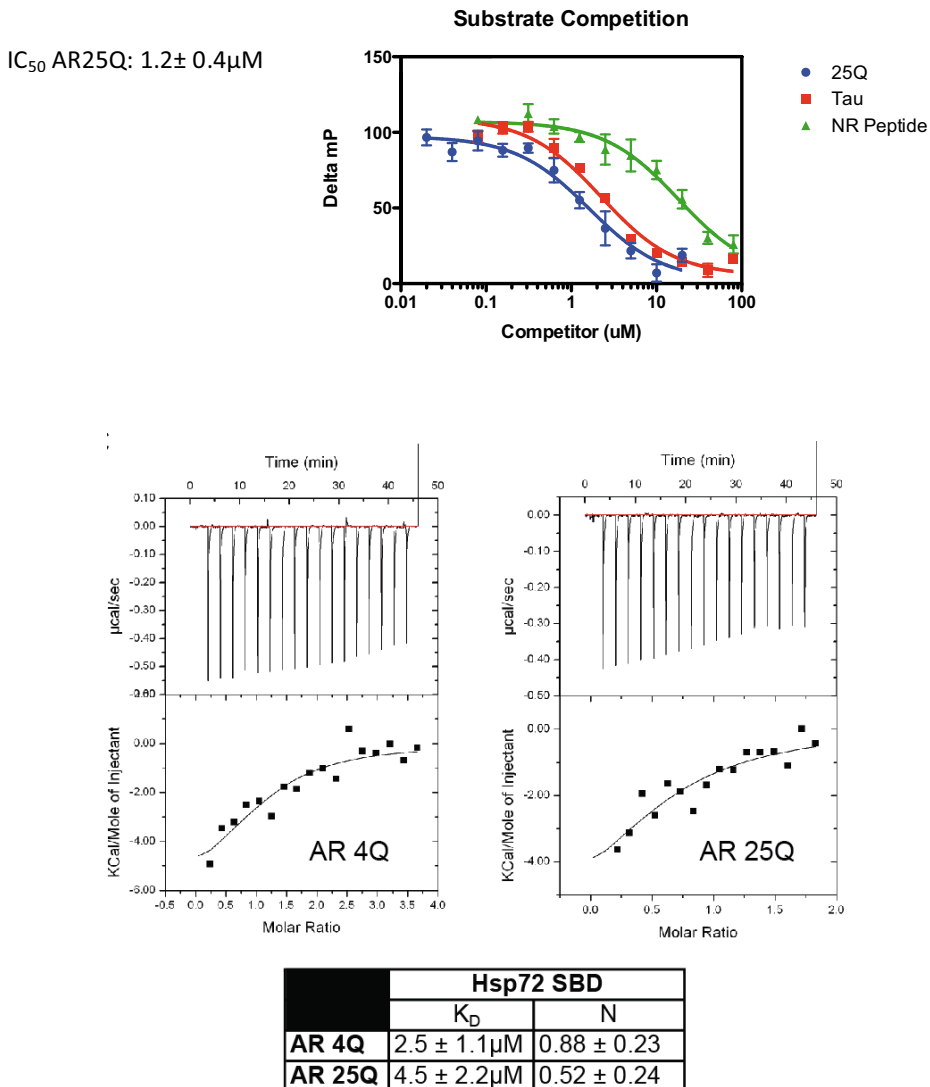


Figure 5-15: AR binds the SBD of Hsp70 and competes with known substrates of Hsp70. (A) Fluorescence Polarization (FP) assay with the SBD of Hsp72 and a substrate tracer LVEALY-FAM. AR4Q and AR25Q compete for binding to Hsp72 with average IC_{50} s of $1.5 \pm 0.04 \mu M$ and $1.7 \pm 0.5 \mu M$ respectively, while a known peptide substrate of Hsp70 (NR) competes with an average IC_{50} of $18 \pm 5 \mu M$. Experiments were performed in triplicate, error is SEM and a representative experiment is shown (B) Isothermal Titration Calorimetry (ITC) experiments show that AR4Q and AR25Q bind the SBD of Hsp72 with μM affinity. Experiments were performed in triplicate, error is SEM and a representative ITC is shown for each.

5.3.7 The ²⁴FQNL²⁸ region is necessary for the interaction of Hsp72 with AR in vivo

Hsp70 is known to preferentially recognize extended peptide segments that are enriched in hydrophobic residues, such as leucine and isoleucine (164). As previously explained, we have identified the binding epitope of AR NTD to Hsp72. Hsp72 binds explicitly to the hydrophobic pocket region of ²⁴FQNL²⁸ and its very proximate flanking region. To validate the mapped region, a mutant of full-length human AR was designed. Our aim was to design a mutant that totally disrupts the interaction between AR and Hsp72. If we hypothesize that Hsp72 binds to AR exclusively by this motif of interest, removing the properties of this hydrophobic pocket by a triple mutant of ²³FQNL²⁷ to ²³AQNAA²⁷ should disrupt the interaction. In collaboration with Claudio di Sanza, who did the experiments and with help of the protein expression facility in IRB, the mutant of full-length construct with a FLAG-tag was transfected to HEK293T cells and proximity ligation assay (PLA) was used to observe the interaction between endogenous Hsp72 and transfected AR- AQNAA.

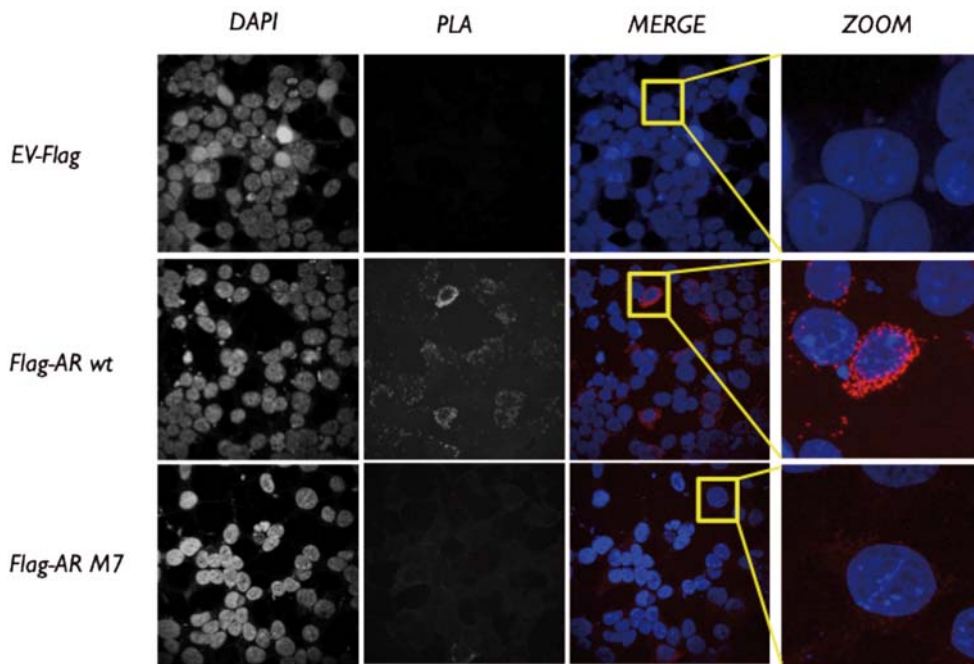


Figure 5-16: Hsp72 interacts with full-length human AR in a sequence specific fashion in HEK293T cells. The cells were transfected with human FLAG-AR construct. Confocal microscopy analysis shows in A) transfected cells with empty vector (EV-FLAG), no positive staining was observed. In B) the transfected cells with FLAG-AR show positive cytoplasmic staining for Hsp72. In C) no positive staining for Hsp72 was seen. The mutant of ²³FQNL²⁷ region to AQNAA, extensively reduces the binding affinity of Hsp72 to FLAG-AR. Images were obtained using a confocal microscope. A representative region from each condition was shown with higher magnification.

As it is illustrated in figure 5-16, confocal microscopy images show that Wt FLAG-AR interacts with Hsp72 and the PLA assay indicated the positive fluorescent immunostaining for Hsp72 in pink. Although, this indicates that AR and Hsp72 interact, it does not have any information about the epitope involved in the interaction. In the lower panel, HEK293 cells were transfected with a triple mutant of AR (AR-AQNAA) from which the hydrophobic residues were removed. An extensive disruption on interaction between AR and Hsp72 is visible by PLA assay (no positive staining). This stresses that the mapped epitope of interaction of full length human AR with Hsp72 is exclusively the ²³FQNL²⁷ region, and no other region on the protein. This is the first demonstration of a precise interaction mapping between AR full-length and the Hsp72.

5.3.8 Over expression of Hsp72 and Hsp40 in C4-2 cells decreases the amount of SDS resistant AR species

For further investigations of the effects of Hsp72 and Hsp40 on polyQ aggregation, C4-2 prostate cancer cells were transfected with two constructs of full length human AR with two different sizes of polyQ (81Q and 25Q). Some of the wells were double transfected with AR and either Hsp40 or Hsp72. As it is shown in figure 5-17, the SDS resistant species of aggregated AR is visible in stacking gel blotted with F39.4.1 mouse monoclonal antibody to the AR N-terminal domain. The cells with over expression of 81Q contain most of the aggregates, while less species in stacking gel seen in 25Q-transfected cells. Our results demonstrate that over expression of Hsp72 in both cells with 81Q and 25Q decrease the positive staining in stacking gel. As well, the overexpression of Hsp40 reduced the aggregated species load (Fig. 5-17).

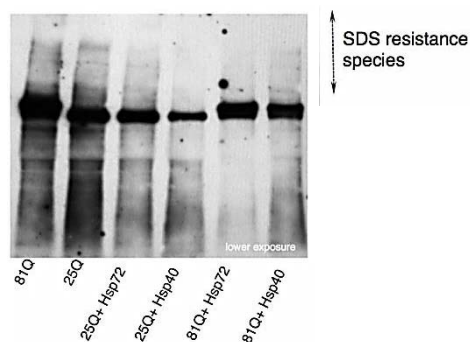


Figure 5-17: Over expression of Hsp72 and Hsp40 in C4-2 cells transfected with full-length human AR decrease the aggregated species of AR in stacking gel. From left to right, cells were transfected with AR-81Q, AR-25Q, 25Q/Hsp72, 25Q/Hsp40, 81Q/Hsp72 and 81Q/Hsp40 and were blotted for AR NTD.

The observation of effects of Hsp72 and Hsp40 on aggregation level of AR were not the key point of this experiment, as it has been previously discussed by other groups (175), but it was important to reproduce the results also in our system, as

our main focus is on the high resolution NMR data. The same cell samples were blotted for Hsp72 antibody. On the left panel, two tissue homogenates of 97Q muscle and spinal cord were also blotted for Hsp72 (Fig. 5-18). These data shows that Hsp72 is present probably together with many truncated variants of AR in tissue homogenates. (It is clear from the rest of the blot that the antibody is very specific for Hsp72 and does not show unspecific binding for other proteins).

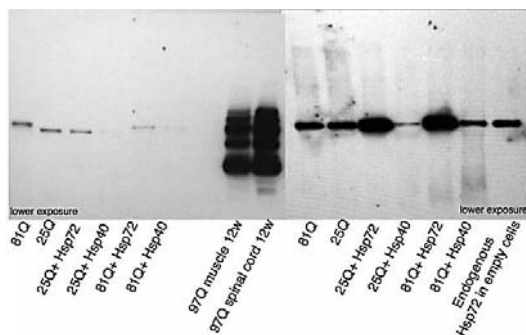


Figure 5-18: Different cells lysates of from cell transfected with 81Q and 25Q were blotted for Hsp72. a) blotted samples for Hsp72 were shown. On the left panel, from left to right, cells were treated with 81Q, 25Q, both 25Q and Hsp72, 25Q and Hsp40, 81Q and Hsp72 and both 81Q and Hsp40 constructs. 97Q muscle and spinal cord samples were also blotted b) blotted samples against Hsp72 were illustrated.

In figure 5-18, the endogenous level of Hsp72 in cells was indicated with neither any transfection of AR nor Hsp72. This blot serves, as complementary information to the figure 5-17 in order to explain that over expression of Hsp is the main reason for the decrease in aggregated species.

5.4. Summary

Due to the importance of Hsps in many neurodegenerative diseases and their multi-functionality which complies with different amyloid proteins, whose nature is not yet well-studied and characterized, we have focused on the identification of interacting epitope of AR with Hsp72 and Hsp40 in this work. These investigations are critical for SBMA research as well as prostate cancer therapeutics strategies (163, 266).

Due to our main interest on role of Hsp72 and Hsp40 in SBMA, in our study, we mapped the interaction between full length Hsp72 and Hsp40 with two constructs from the N-terminal domain of AR. To determine a precise binding region for both Hsp, the binding sites were mapped by NMR. We acquired HSQC spectra of AR-25Q and AR-4Q in absence and presence of different concentrations of Hsp40 and Hsp72 and checked the signal intensity of each individual peaks. This is the most precise way of mapping a protein-protein interaction. We found all segments of specific binding of AR with Hsp40 and Hsp72. Surprisingly, for Hsp72, the main binding motif was not one of the predicted residues. Our studies have show that Hsp72 mainly binds to ²⁴FQNL²⁸ regions and protect this hydrophobic pocket

from forming intermolecular interactions, which may lead to oligomerization and aggregation of AR (see Figs. 4-9 and 4-10).

While Hsp72 exclusively interacts with AR-25Q and AR-4Q by ²⁴FQNLF²⁸ region, Hsp40 shares the same motif with Hsp72, though with less specificity and as well binds to two other hydrophobic segments of 4Q and 25Q proteins in ⁵⁵LLLLQQQ⁶² and ¹⁰⁶GPTGYLVL¹¹⁴ regions.

Interestingly, our complementary studies in cells have shown that for Hsp72, the ²⁴FQNLF²⁸ region is exclusively the binding region in full-length AR and therefore disruptions of this hydrophobic binding pocket by mutation completely eliminate the interaction.

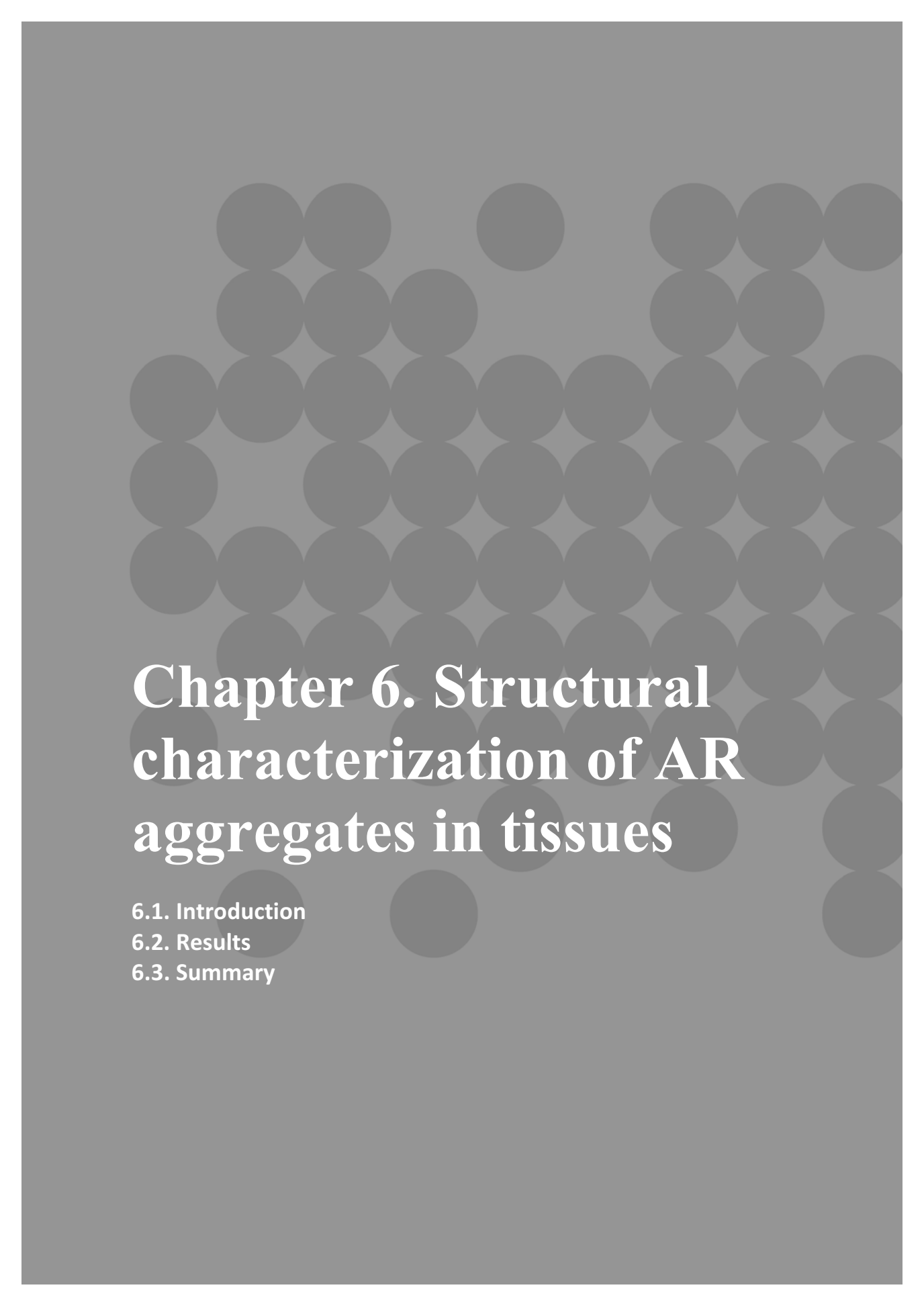
NMR experiments in which 4Q and 25Q were incubated with two of such potential interactors, supposed to illustrate the tentative presence of positive or negative cooperativity *i.e.* whether the presence of one interactor affected the interaction of 4Q with a second one. However, our results indicated no clear cooperativity in presence of both Hsp40 and Hsp72.

Considering the ATPase activity of Hsp72, carrying out experiments in the presence of ATP, we studied whether this compound influenced the interaction of 4Q and 25Q with Hsp40 and Hsp72 as well as the degree of cooperativity of the interaction. We observed few small changes in peak intensities in both 4Q and 25Q, in presence of ATP. More investigation is needed to elucidate the role of ATP *in vitro* in these experimental conditions.

There are already many studies on the importance of Hsp in SBMA pathology and therapeutics (175, 176); the mapping data on the interaction of AR with Hsp72 and Hsp40 is now fully available. Thus, we propose that Hsp interaction-modulators (39) with their potential use for drug screening and therapeutics can be easily studied using our system.

Our findings highlight the therapeutic potential of allosteric regulators of Hsp70 and Hsp40 and provide new insights into the role of the chaperone machinery in protein quality control (267).

This is to note that the therapeutic strategies are not only beneficial for SBMA patients, but also in prostate cancer studies (244, 264). Now is the correct time to design small molecules that could strengthen or weaken the interactions between Hsps and amyloid proteins (Hsp-modulators). Thus, we suggest that targeting Hsp70 and Hsp40 interaction with AR, not only with elongated polyQ but also in Wt variants, can lead to new therapeutic approaches for SBMA and prostate cancer (179, 264, 266, 268, 269).

The background of the page is a solid dark gray color. Overlaid on this background is a pattern of numerous overlapping circles of varying shades of gray, creating a textured, cellular appearance. The circles are arranged in a somewhat irregular grid, with some overlapping significantly more than others.

Chapter 6. Structural characterization of AR aggregates in tissues

6.1. Introduction

6.2. Results

6.3. Summary

Chapter 6. Structural characterization of AR aggregates in tissues

6.1. Introduction

To characterize the AR aggregates that form *in vivo*, we brought the colony of transgenic mouse models of SBMA generated in Prof. Sobue laboratory at University of Nagoya, carrying full-length AR with either 97Q or 24Q (33). Three of the five lines showed progressive muscular atrophy and weakness as well as diffuse nuclear staining and nuclear inclusions consisting of mutant AR. These phenotypes were markedly pronounced in male transgenic mice and rescued by castration (94). Female transgenic mice showed only mild phenotype, a few manifestations, which increased by aging of the females, however markedly deteriorated with testosterone administration. This SBMA mouse model can be considered as the best one, as the animals either show no phenotype and look identical to Wt mice when they comprise the 24Q AR gene or show a significant SBMA pathology when the polyQ length extends to 97 (33, 94). Therefore, the colony was generated from embryos in Barcelona, as previously explained in Materials and Methods.

To date, many studies on transgenic mouse models of SBMA have characterized the pathology and histology of the disease (33, 82, 198). Other studies have investigated different approaches for developing therapeutic strategies to tackle the disease or improve the life conditions of the patients (270). From our point of view, what is still missing and might be the possible reason for not yet being close to a cure or preventive medication is that the molecular basis of the disease has not been completely studied.

One of the problems in studying neurodegenerative diseases is that most of the time scientists look for similarities between different diseases' manifestations and use the best-generalizing approach to study different diseases in very similar way. It seems that this happens also in the polyQ diseases field of research. Therefore, in this work, we decided to go one step back and study the missing points in pathology of SBMA, without accepting it as a neuromuscular disease with central nervous system as the main starting point of the disease.

For this, we performed a thorough biophysical and biochemical characterization of tissue homogenates of SBMA transgenic mice and suggested a possible explanation for the mechanisms of toxicity of AR aggregates in SBMA pathology.

6.2. Results

6.2.1 Histological and biochemical characterization of 97Q and 24Q tissue homogenates

In this SBMA transgenic animal model, the overexpressed human AR with 97Q elongated tract aggregates and forms intranuclear inclusions in spinal cord and skeletal muscle. These aggregates are detectable with specific antibody (IC2). In 24Q animals, the overexpressed human AR with only 24 Gln does not aggregate and the animals do not show any SBMA-like pathology.

In figure 6-1 and 6-2, immunohistochemical analyses of muscle and spinal cord tissues in 97Q and 24Q are compared. For immunohistochemistry analysis of paraffin-embedded tissues, the 5µm cuts of muscle and spinal cord tissues were fixed and prepared as previously described and finally were blotted by using the IC2 antibody. Staining results of aggregated species with higher number polyQ over 38 revealed that (1) 24Q animals have no positive staining neither in muscle cells nor in spinal cord and (2) the number of aggregated species detected in muscles of 97Q animals is significantly higher than in their spinal cord.

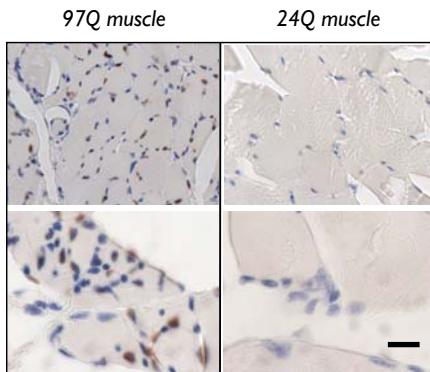
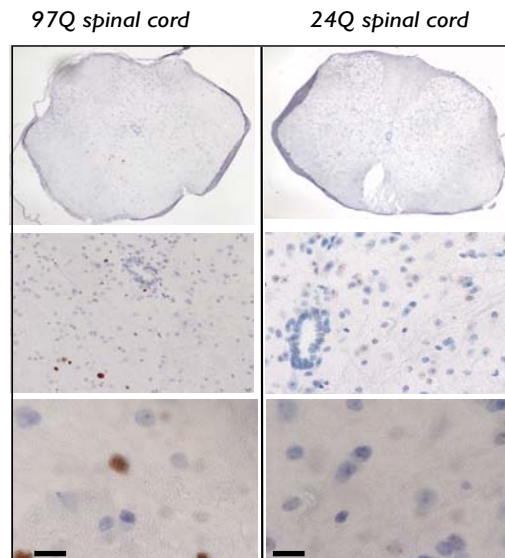


Figure 6-1: Immuno-histochemical analysis of muscle tissue of 97Q and 24Q transgenic mice of SBMA by mouse monoclonal IC2 antibody (1:50000) specific for elongated polyQ over 38. In 97Q males diffuse nuclear staining presents while in 24Q animals neither in muscle nor in spinal cord tissues no positive staining was observed. Scale bar= 20µm

Figure 6-2: Immuno-histochemical analysis of spinal cord tissue of 97Q and 24Q transgenic mice of SBMA by mouse monoclonal IC2 antibody (1:50000) specific for elongated polyQ over 38. In 97Q males diffuse nuclear staining presents while in 24Q animals neither in muscle nor in spinal cord tissues no positive staining was observed. Scale bar= 20µm



The immunohistological data were accompanied by biochemical characterization of tissue homogenates of 97Q and 24Q animals (Fig. 6-3).

WB analyses of total homogenates from muscle and spinal cord tissues of male AR-97Q and AR-24Q (12-week-old) mice were immunolabeled by an antibody (H280) for the AR N-terminal domain. Mouse AR was hardly but still detectable. Most of the lower bands in transgenic mice represent truncated human AR (Fig. 6-3 and 6-4).

As indicated in figure 6-4, similar to the immunohistological staining results, aggregated AR is present in much higher quantity in muscles compared to spinal cord. Protein species in stacking gel reveal the presence of higher molecular weight aggregates of AR in 97Q mouse tissues, while in 24Q animals no stacking material was detected.

A single band of AR monomer consistent with 24 and 97 glutamines and truncated fragments of mutant AR in all symptomatic lines is shown in figure 6-3 and figure 6-4. This is another representative WB of total tissue homogenates of three different tissues extracted from AR-97Q and AR-24Q.

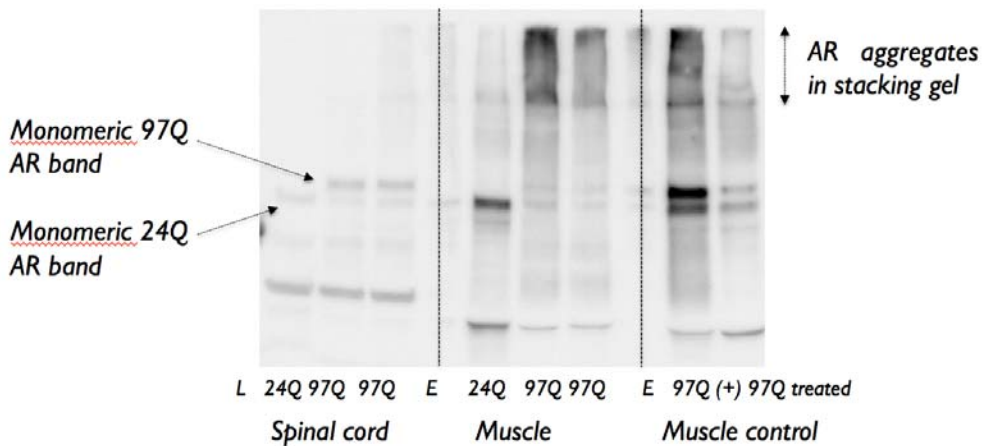


Figure 6-3: WB analysis of AR in tissue homogenates of AR-24Q and AR-97Q animals. Spinal cord and muscle tissue from 24Q male mouse and spinal cord and muscle samples from 97Q mice were analyzed and were immunoblotted by H280 antibody. The monomeric bands of A- 97Q and AR-24Q \approx 110kDa are indicated. The lower bands present in AR 97Q animals (at the same level as 24Q- AR band) indicate the endogenous mouse AR. AR aggregates in stacking gel are indicated.

Due to problems in preparing the tissue homogenate to identify the aggregated species at the beginning, this WB has been repeated overly, until the reproducible protocol was achieved.

In figure 6-4, three different tissues (Spinal cord, brain stem and muscles) from 97Q and 24Q animals were blotted against human AR. It was shown that Central nervous system (CNS) comprised less aggregated species compared to muscle tissue, not only in 97Q, but also in 24Q. 80 μ g of muscle tissue and 120 μ g of CNS

tissues were immunoblotted, however the difference between aggregate quantities is still significant.

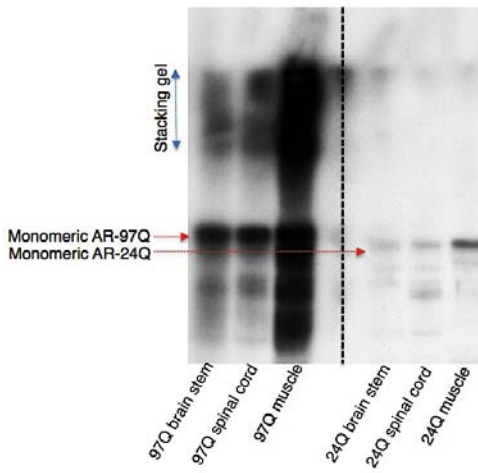


Figure 6-4: WB analysis of AR in tissue homogenates of AR-24Q and AR-97Q animals. 80µg muscle tissue and 120µg spinal cord and brain stem tissue from each group of animals were prepared and immunoblotted by H280 N-terminal domain antibody. Monomeric AR of both 97Q and 24Q are indicated. A blue arrow marks stacking gel containing the SDS resistance material.

Our observation on different quantities of proteins in spinal cord and muscle as well the difference between 97Q and 24Q in immunohistochemical analysis, hinted us to check the protein expression level by measuring the relative total AR mRNA level in transgenic animal tissues. In figure 6-5, expression levels of AR in muscle and spinal cord in both transgenic animals of SBMA are compared.

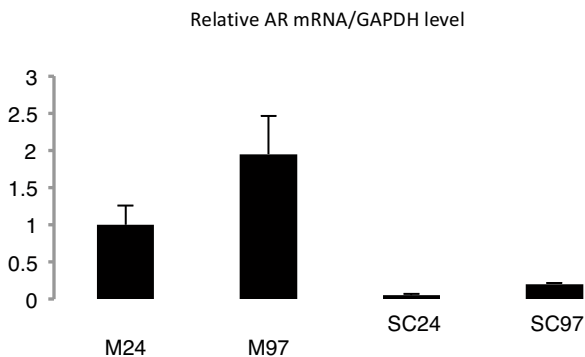


Figure 6-5: Expression levels of AR in muscle and spinal cord of 12 weeks old 97Q and 24Q transgenic animals of SBMA. Quantification of AR mRNA levels relative to GAPDH by RT-PCR analysis shows a significant difference between the level of AR in spinal cord and muscle in both groups. n=3

The expression level of AR was measured relative to GAPDH reporter gene in 4 tissues. Comparing the biochemical analysis by WB and protein expression level by RT-PCR showed that the increase in protein level and mRNA level have similar trends. This data strongly suggests that the amounts of aggregated species are different in various tissues and this is mainly due to differences in the protein expression level.

6.2.2 PolyQ AR aggregates are SDS resistant, but not Proteinase-K (PK) resistant

Previous observations have shown that polyQ aggregates, independent of their amino acid composition, are SDS resistant. PolyQ aggregates prepared *in vitro* from Htt exon-1 by Ronald Wetzel laboratory, Htt aggregates isolated from tissue homogenates of HD patients and AR in tissue homogenates of SBMA transgenic animals all share one similarity in their biochemical properties on WB. All of these aggregated species are SDS resistant. This explains the presence of positive antibody staining in the stacking gel in WB. SDS is usually used to denature the proteins and make a linear protein structure that can easily run on the gel. In case of polyQ aggregates, the extensive strong intermolecular interaction between monomers of proteins avoid SDS to solubilize them and therefore, they stay as aggregates in the stacking part of the gels. The SDS resistant property belongs only to polyQ proteins and none of the other amyloid proteins, while in other proteins like A β aggregates in AD or prion proteins; the toxic aggregates show Proteinase K (PK) resistance. Therefore, it was interesting to check the properties of AR aggregates by PK digestion.

To further characterize the properties of tissue homogenates of SBMA transgenic animals, we digested 10%w/v total muscle tissue homogenates of 24Q and 97Q and added 50 μ M PK and digested them for 1h at 37°C. To have a closer look into general properties of AR aggregates, two N-terminally truncated AR recombinant constructs with 51Q and 25Q aggregates were prepared and used as controls. The respective samples were also digested with different concentrations of PK.

Total tissue homogenates of 24Q and 97Q muscle tissue homogenate digested with/without PK (50 μ M)

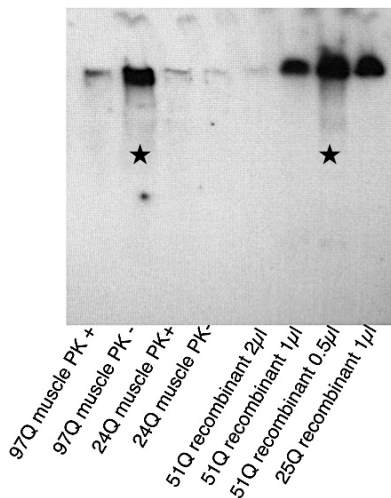


Figure 6-6: PolyQ AR aggregates are SDS resistant but not Proteinase- K (PK) resistant. For 1h at 37°C, total muscle tissue homogenates from 24Q and 97Q mice and recombinant 25Q and 51Q N-terminal construct of AR were digested by PK and immunoblotted by H280. SDS resistant species in stacking gel indicate only the presence of these species in 97Q muscle tissue and in absence of PK, while in 24Q tissue homogenates neither SDS resistant nor PK resistance species identified. 51Q recombinant proteins are PK sensitive and show a concentration dependent behavior.

As it was shown in figure 6-6, polyQ aggregates are sensitive to PK digestion and even the SDS resistance species did not withstand the PK treatment. PolyQ aggregates from 51Q recombinant protein showed concentration dependent PK sensitivity.

6.2.3 PolyQ AR aggregates in muscle comprise truncated N-terminal AR; spinal cord contain mainly C-terminal species

In order to further characterize the tissue homogenates of 97Q SBMA mice, two different antibodies staining on WB biochemical analysis were performed. Our previous results showed that H280 antibody; specific for the epitope of 91-370 of human AR-NTD, stained the SDS-resistant material in stacking gels. This indicates that these materials comprise the N-terminal part of AR, either the whole or a truncated variety of it. It is not yet clear in the field, similar to what is the case in HD, where Htt exon-1 is the main involved species in aggregation, if only N-terminal fragment of AR is critical for SBMA pathology or the full length protein is mandatory for toxicity. Therefore, we investigated this using a specific antibody for C-terminal part of the protein (Fig. 6-7).

Comparison of N and C Terminal Antibody H280 for N-terminal (a) and C19 for C-terminal (b)

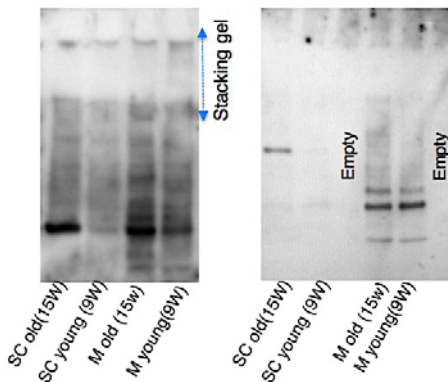


Figure 6-7: Aggregates in stacking gel comprise the N-terminal domain of AR but not the C-terminal. 10%w/v total tissue homogenate of muscle and spinal cord of young and old SBMA transgenic mice were prepared and blotted by H280 (left) and C-19 (right). The SDS resistant material in muscle tissues is only positive and detectable by N-terminal specific antibody.

The biochemical analysis showed that the AR aggregates in the stacking gel are not positive for C-19 staining (Fig. 6-7-right), whereas they are positive for H280 (Fig. 6-7-left). This clearly indicates that the aggregates do not comprise the C-terminal part of AR. We characterized these features of aggregates in more optimized conditions and with more 97Q transgenic animal's muscle and spinal cord tissues. Figure 6-8, shows the biochemical analysis of spinal cord (SC) and muscle (M) tissue homogenates of four different 97Q SBMA mice. In the upper panel, C-19 blotted samples and in the lower panel the same samples only blotted by H280, were shown.

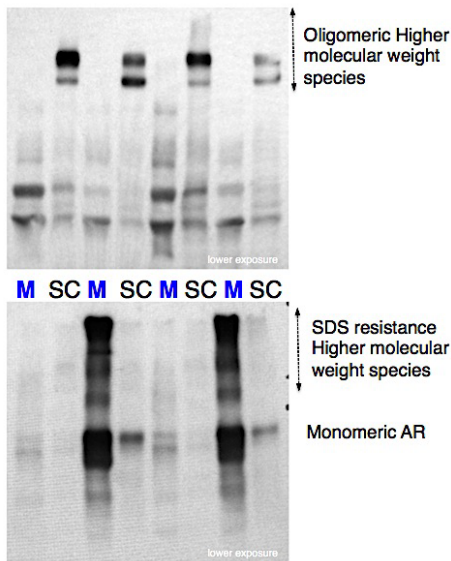


Figure 6-8: Aggregates from muscle tissues in stacking gel comprise the N-terminal domain of AR whereas C-terminal positively high molecular weight species were detected only in spinal cord. 10%w/v total tissue homogenate of muscle and spinal cord of four SBMA transgenic mice were prepared and blotted by C-19 (upper panel) and H280 (lower panel) antibodies. The SDS resistant material in muscle tissues is only positive for N-terminal antibody while spinal cord tissue comprises higher molecular oligomeric species positive for C-19 antibody.

Four different muscle and spinal cord total tissue homogenates from middle age and old SBMA 97Q mice were characterized by C-19 and H280 antibodies. The lower panel indicates the SDS resistant species positive for H280 N-terminal domain antibody as expected. Interestingly, in contrast, we observed (upper panel) that in all groups of spinal cord tissues, C-19 detects higher molecular weight oligomeric species exclusively, whereas no positive staining was detected in muscles. This may explain the main yet unclear differences between the aggregates in muscle and spinal cord.

To summarize these data, it is to note that (1) AR aggregates in muscles differ from spinal cord, not only in their structures as it was observed by AFM (in the following- Fig. 6-13), but also in their biochemical properties (fig. 6-8). (2)- Muscle tissue aggregates comprise AR-N terminal domain, whereas there is no proof of presence of C-terminal (LBD) (no positive staining with C19). Nevertheless, interestingly, (3) spinal cord aggregates, show much less SDS resistant material and there is clear proof of presence of positive C-terminal staining, which refers to high molecular weight oligomers.

6.2.4 Seprion ligand specifically captures AR aggregates in total tissue homogenates of SBMA transgenic animals and provides a highly quantitative assay

The Seprion ligand has been previously used to isolate and quantify aggregated forms of prion protein (271). It is a high molecular weight polymeric ligand that consists of repeating charged and hydrophobic chemical groups that interact with similar repeating groups that occur on aggregated proteins (previously described in M&M). The selectivity of aggregated proteins is based on the arrangement of large numbers of polar and hydrophobic regions in aggregating protein, which

occur to a much lesser extent in a single unaggregated protein molecule. We used the Seprion ligand. It is developed as an ELISA-based plate assay, by which aggregated AR from tissue lysate were captured and subsequently detected by immunoprobng against H280 antibody specific for N-terminal of AR. The amount of aggregate load in brain stem, spinal cord and muscle tissues of 14 weeks 97Q and 24Q SBMA transgenic mice were quantified (Fig. 6-9). Monomeric polyQ were prepared *in vitro* as a control. Immunoblotting controls of primary and secondary antibodies were also analyzed.

Our results indicated that the Seprion kit detects exclusively AR aggregates and none of the monomeric form of AR. Therefore; there was no positive ELISA signal from fluorescence absorbance in 450nm for polyQ monomers, neither in control wells for antibodies nor in 24Q animals (Fig. 6-9). The samples from 97Q animals showed positive signal in ELISA kit. The signal intensity was very high in muscle samples; much less in spinal cord and very little signal was detected in brain stem (not significant).

Therefore the Seprion ELISA might provide a highly sensitive method of quantifying the amount of aggregate loads in tissues from not only the mouse model of SBMA but also other polyQ diseases. This assay has great potential to be used as a readout measuring system, either in preclinical drug trials to test the effects of potential therapeutic interventions or in transgenic animal generation, as it provides us with very quick data regarding the aggregates load. And finally, this is to emphasize that this is the first well-optimized system used for polyQ aggregation detection of *ex vivo* material.

Finally, our results illustrate that among all different tissues, muscle tissue has significantly much higher loads of aggregated species. It is interesting to see that, similar to IHC data, AR aggregate loads in muscle keeping the same trend compare to Spinal cord.

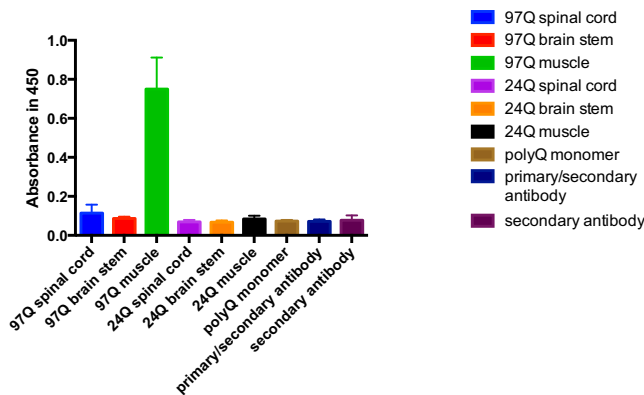


Figure 6-9: Seprion ligand quantification of aggregate load in tissues from SBMA mouse model. Quantification of aggregate loads in brain region and muscle shows a significant difference in the amount of aggregates in 97Q animals compared to 24Q animals. 97Q muscle tissue aggregate loads from three mice (in three independent experiments) are significantly higher than 24Q muscle tissue.

ANOVA summary	
F	51.75
P value	< 0.0001
P value summary	****
Are differences among means statistically significant? (P < 0.05)	Yes
R square	0.9474

6.2.5 Seprion ligand-coated magnetic beads capture AR aggregated species for biophysical characterization

We employed the same Seprion ligands developed as magnetic beads in order to capture the aggregated species and pull down the aggregates from tissue extracts. In spite of the similarity of Seprion ligand and convectional pull down or immunoprecipitation, Seprion kit provided us with many advantages over conventional methods. With Seprion magnetic beads, we avoided using any antibodies. In most of amyloid proteins, as the binding epitopes of amyloid proteins are hidden in the aggregates core, it is challenging to use antibody to capture the protein of interest. Therefore, an antibody free system could be highly favored. Moreover, due to the same reason, the samples prepared by Seprion ligand are free from any antibodies contamination (use for immunoprecipitation), which are usually troublesome in any mass spectroscopic analysis.

We employed SDS-PAGE and WB analysis to investigate the aggregated, SDS resistant materials that were captured from muscle tissues lysate and eluted from the Seprion ligand. Our results indicate (Fig. 6-10) that eluted material from Seprion ligand from transgenic mouse tissue contains only aggregated species, which remain in stacking gel. No monomeric form of protein was detected.

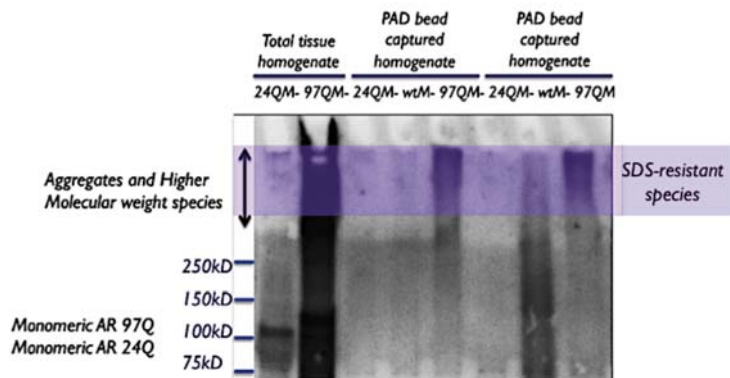


Figure 6-10: SDS-PAGE and immunoblotting of Seprion bead-captured aggregates from Wt and SBMA transgenic mouse muscle tissue. Seprion ligand bead-captured materials were fractionated by 5-20% SDS-PAGE alongside the corresponding mouse lysates. Blots were immunoprobed with H280 antibody. High molecular bands detected by H280 that stay in the stacking gel are indicated in violet. Indicate fragments that resolve at a size similar to monomeric AR, are differentially recognized by the similar AR antibody in total tissue homogenates. The order of the samples in gel: 24Q muscle, 97Q muscle, and 24Q-Wt-97Q captured aggregates in SDS, 24Q-Wt-97Q captured aggregates in SDS and heating condition.

In figure 6-10, total tissue homogenates of 24Q and 97Q muscles were blotted beside the Seprion ligand captured material from 24Q, Wt and 97Q muscles. Aggregates from two mice from each group were analyzed and only in 97Q mice muscle tissue SDS resistance aggregated species were detectable (identified by

color). It took some months to have this capturing detection system optimized, but as soon as we found the condition, it was reproducible, and therefore I enclosed also two more blots of Wt, 24Q and 97Q animals, which show all the same pattern in detected material by Seprion beads (Fig. 6-11).

Finding a method to capture and quantify the aggregated species led us to further characterization of the captured aggregated species. We isolated the aggregated species from tissues and visualize them by using AFM and TEM in a time-lapsed fashion. We were interested in correlating the disease manifestations with the presence of toxic species that are responsible for the disease pathology.

Two Seprion Trial stained with H280

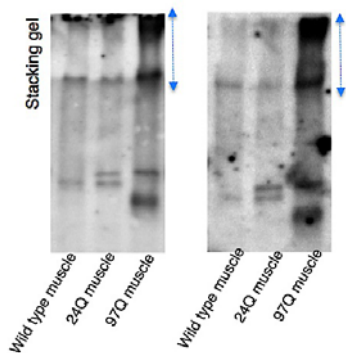


Figure 6-11: SDS-PAGE and immunoblotting of Seprion bead-captured aggregates from Wt, 24Q and 97Q animals' muscle tissue. Both panels show 2 mice samples from all three groups of animals. Only positive SDS resistant aggregates were found in 97Q muscle tissue.

Therefore, different ages of the animals' tissue were compared (Fig. 6-12). We found that as the animal ages, the aggregated species also increase in load. This interesting observation prompted us to explore further the structural properties of tissue homogenates by capturing the aggregates and study them by high-resolution methods with the hope of explaining the disease mechanism.

Total tissue homogenates of various 97Q animals with different ages have been compared. Staining with H280 Ab

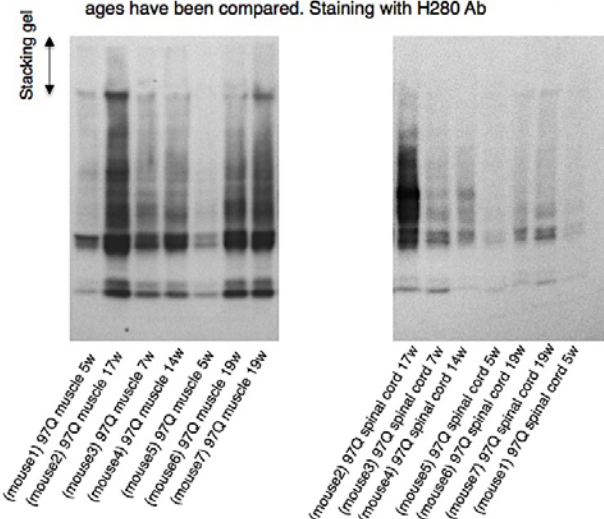


Figure 6-12: Total tissue homogenates of various ages of 97Q animals show an age dependent aggregate loads in stacking gel. Muscle tissues (left) comprise more aggregated species compare to spinal cord (right). Samples were blotted by H280 specific for N-terminal domain of AR.

6.2.6 Fibrillar aggregates species from SBMA transgenic animals muscle tissue have identical structural properties as those generated from N-terminal truncated AR construct *in vitro*

We applied AFM and TEM to provide more information on the structural properties of aggregates in muscle and spinal cord of SBMA 97Q mouse. The aggregates were captured by the Seprion ligand as was previously explained. As we were mainly looking for SDS-resistance species of AR aggregates in tissue, we eluted the aggregates from beads by SDS. We optimized this condition to be certain that all large species are released from beads.

Three different groups of animals in three age groups of young animals (5-6 w old), middle age (9-10 w old) and old animals (14-17 w old) were studied. Our strategy to group the animal was based on the manifestations of the disease. The animals, which show no phenotype for the disease and look identical to Wt and 24Q mice of the same age, belong to the “young” or pre-onset group. In the middle age animal group, the first signs of the disease are visible, but they still have cage activity and can easily access their food, however they have muscle mass loss (Pre-symptomatic). The animals with no cage activity belong to the “old” group; these are mostly end stage animals with clear muscle atrophy.

Age correlation studies in figure 6-13 indicated that (1) surprisingly; the capture and eluted aggregated species in muscle and spinal cord tissue obviously differ (Fig. 6-13 and 6-15). In muscle tissue, mainly fibrillar-aggregated species were found, whereas in spinal cord more of globular pre-fibrillar aggregates were captured. (2) In young animals, in both muscle and spinal cord, only oligomeric species of AR aggregates were found. These species, exclusively in muscles, evolved along the time as the animal ages, not only the size of these species, but also number of these fibrils increased during aging and progression of the disease. In contrast, no significant changes were observed neither in size nor in quantity of aggregates in spinal cord.

In the group of old animals the captured species were mainly fibrillar species with an average size of 1 μ m in length. Interestingly, the fibrillar species appeared to be identical to those that had been previously generated *in vitro* by AR N-terminal truncated constructs (Fig. 6-14). This information was another hints to the role of N-terminal AR variant in AR aggregation *in vivo*.

As it was previously discussed, our biochemical analysis of muscle tissue homogenates of 97Q animals by WB validates that the antibody specific for N-terminal domain of AR did only detect the SDS-resistance material in stacking gel, whereas the C-terminal specific antibody did not recognize any epitope for binding (Fig. 6-7). Therefore, we hypothesize that the main species in fibrillar AR aggregates in muscle comprise mainly the N-terminal the Caspase-cleaved truncated AR and not the full-length. As it is shown in figure 6-14, the structural features of aggregates in 97Q muscle captured aggregates are very similar to

recombinant protein from N-terminal truncation of AR construct (51Q). This is another strong suggestion that the aggregated species do not comprise full-length AR.

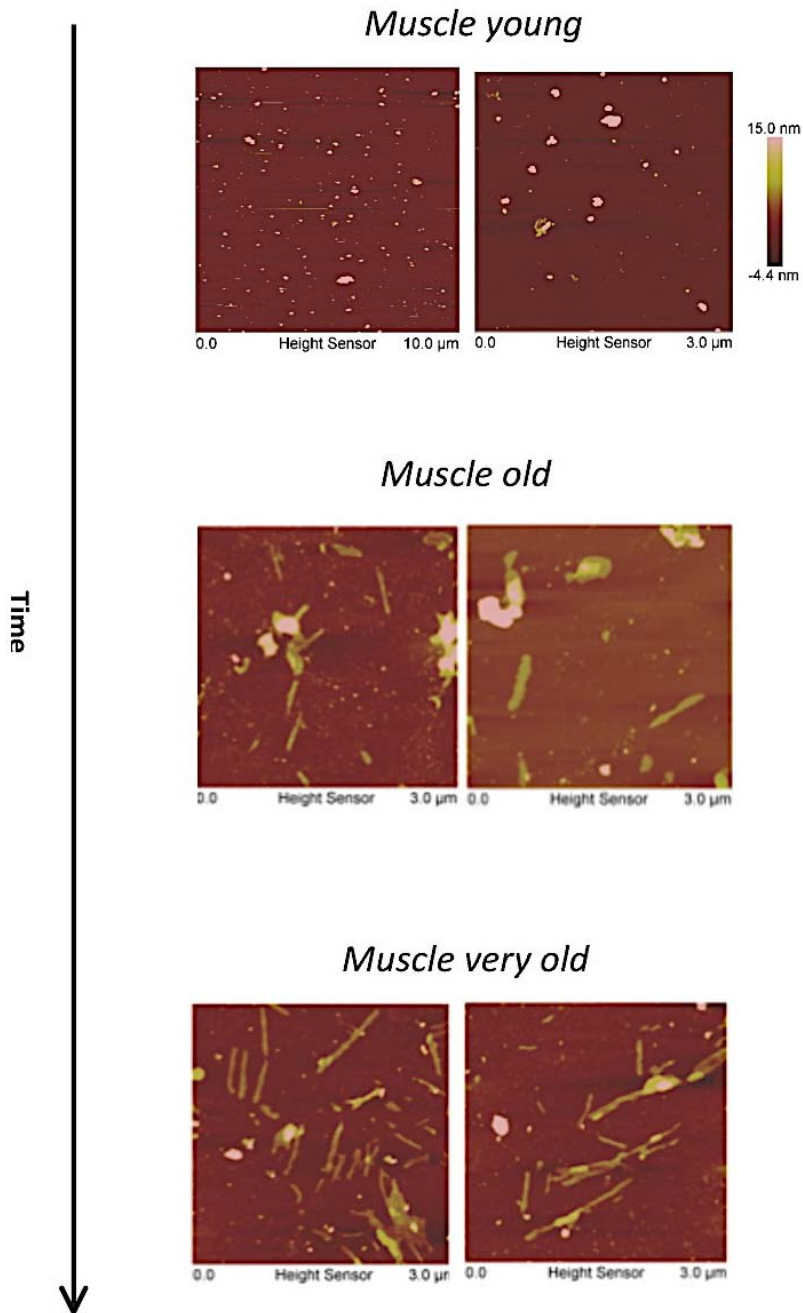


Figure 6-13. AFM analysis of nanometer fibrillar aggregates from 97Q male SBMA mouse muscle. AFM of Seprion ligand- captured material from 97Q muscles at 3 different age groups indicated that there is a correlation between the numbers of fibrillar species and their length with age.

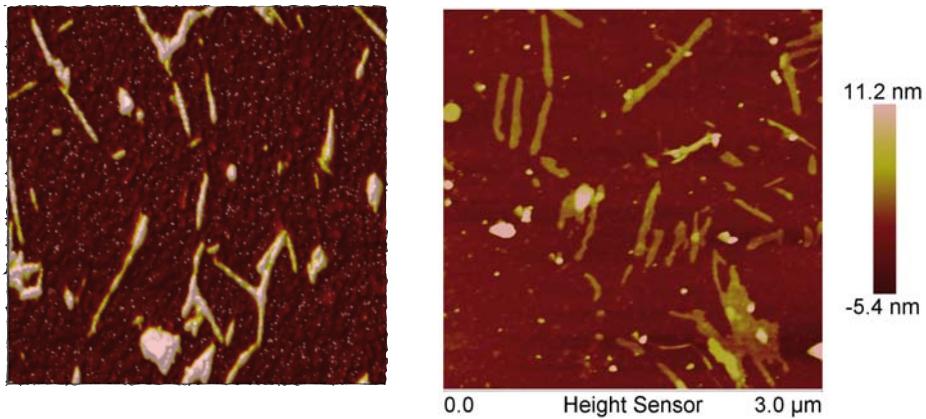


Figure 6-14: AFM image comparison of 51Q N-terminal domain AR recombinant protein (left) and AR captured aggregates from 97Q mouse muscle tissue homogenate (right) shows a close similarity in the appearance (form and size) of ex vivo aggregated species to in vitro N-terminally truncated variant of AR. The fibrillar species are in size order of 300nm to 1μm long.

In collaboration with Dr. Tobias Jochum, we carried out a detailed quantification of AFM data to have a complete characterization of all structural variants of AR in muscle and spinal cord tissue homogenates of 97Q animals. Three groups of aggregates were quantified in both tissue homogenates, amorphous, globular and fibrillar species (Fig. 6-16). In all three groups, we quantified a higher total number of species in muscle tissue compared to spinal cord. It was already predicted from our data that AR present and aggregates in much higher quantity in muscle cells compared to spinal cord and therefore, it was obvious that the captured species would be also in large amounts in muscles. Interestingly, the globular and amorphous aggregates percentage is almost constant in three different ages of animals in muscle tissue, while the fibrillar species increase significantly as the animals get older (Fig. 6-16). Our results in figure 6-17 indicated that total amount of SDS resistance species of AR aggregates in muscle tissue increase, as the animals get older. Moreover, the amount of aggregated species in old and symptomatic mice with no cage activity is significantly higher than other groups. This data is a clear validation for the importance and the role of AR in muscle tissue in disease features and pathology. Based on our observations, we suggest that the main tissue involved in the disease is muscle and there is a clear correlation between the number and length of aggregated species and phenotype in SBMA 97Q animals.

This is the first report at molecular level resolution on the importance of muscle tissue in SBMA pathology (90). These observations open new windows to the molecular mechanisms of the disease and pathology. We hope that our high-resolution data provides the field with better definition for SBMA; muscle degeneration prior to motor neuron degeneration.

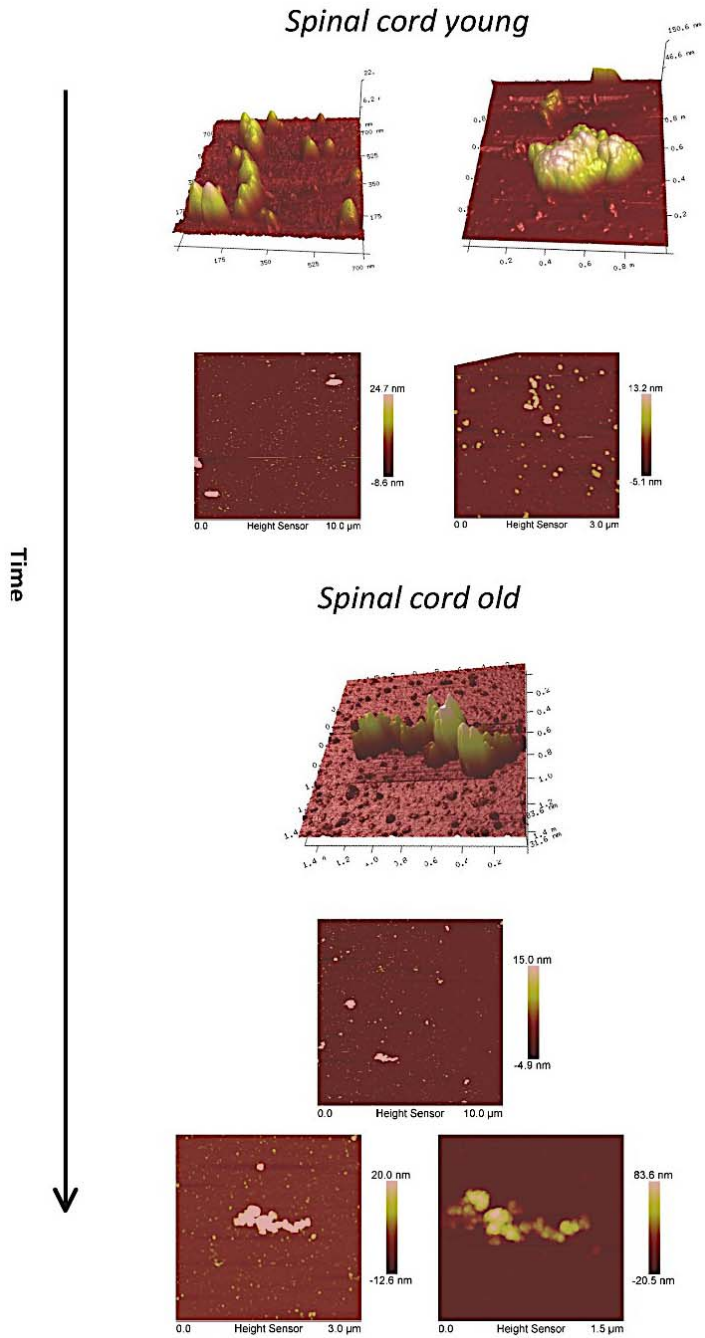


Figure 6-15: AFM analysis of nanometer granular and pre-fibrillar aggregates from 97Q male SBMA mouse spinal cord. AFM of Seprion ligand- captured material from 97Q spinal cord at 3 different age groups indicated that there is a correlation between the numbers of pre-fibrillar species and their length with age.

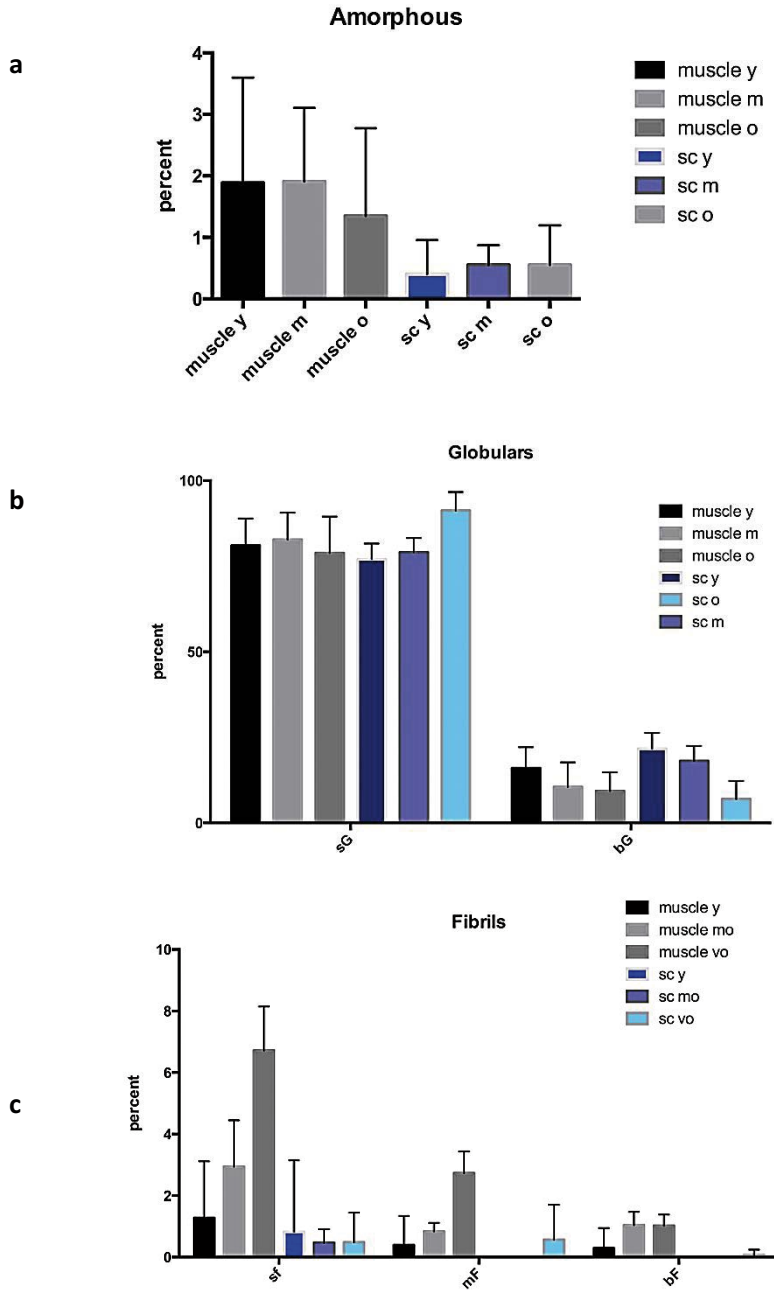


Figure 6-16: Septrin ligand quantification of aggregates types from AFM data. Three groups of observed aggregates in AFM data (amorphous (a), Globular (b) and fibrillar species (c)) were quantified. Three groups of animal's muscle tissue (m) and spinal cord (sc) were compared, young animals (y- pre-symptomatic), middle age animals (m- beginning of onset of the disease) and old animals (o- no cage activity) were compared.

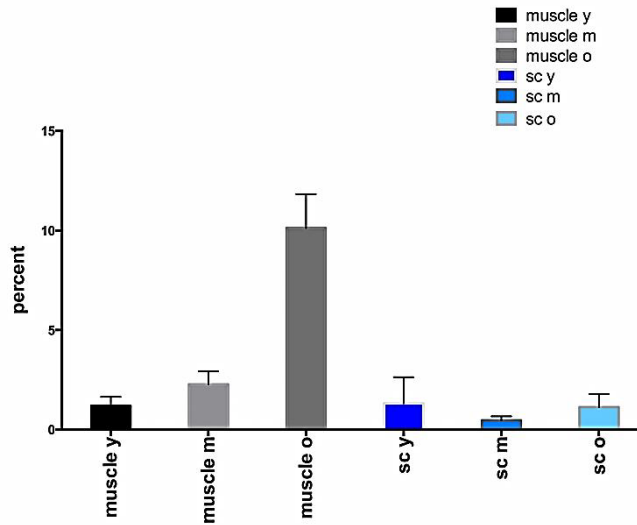


Figure 6-17: Total percentage of Seprion ligand captured aggregates imaged and analyzed by AFM, shows a significant increase in the SDS resistant AR species in Seprion captured material in the group of old animals. Each column shows the total percentage of captured aggregates from Seprion beads. Six groups of muscle young (y), muscle middle age (m), muscle old (o), spinal cord young (sc y), spinal cord middle age (sc m) and spinal cord old (sc o) were compared.

The quantified data summarize that small, middle size and large fibrils exist in higher quantity in mice with no cage activity compared to rest of the groups. As it was shown in figure 6-17, the total percentage of captured aggregates is also significantly higher in muscle tissue of an end stage mouse. The progression is slower at the beginning of the disease's onset, while it increases exponentially in the later stages of the disease.

To summarize this data and the biochemical characterizations of muscle and spinal cord, we hypothesize that for AR to make fibrillar aggregates in muscles, as it is shown in our AFM data (Fig. 6-13), it might truncate and comprise only the N-terminal fragments, while AR aggregates in spinal cord contain the C-terminal domain and might not go under any modifications which make the protein prone to aggregation, unless they change in later stages of the disease. This may explain the absence of fibrillar species in the spinal cord.

6.2.7 AR aggregates have well-defined fibrillar structures in muscle tissue of 97Q males of SBMA

Our AFM results provided us with more evidence in the importance of the fibrillar species in muscle tissues of 97Q mice; therefore the *in vivo* fibrils were also analyzed by TEM. Due to the fact that AFM images in tapping mode only reflect the structural properties of specimens on the mica's surface, it was complementary for us to visualize the structures of the fibrils by directly TEM. AR

containing tissue homogenates were prepared and aggregates were captured by Seprion kit as was previously explained. The aggregates were spotted on copper grids and false positively stained. As it is shown in figure 6-19, the captured aggregates from *in vivo* material were visualized for the first time by TEM. Clear fibrillar species, similar to our observations in AFM, were detected in muscles tissue homogenates.

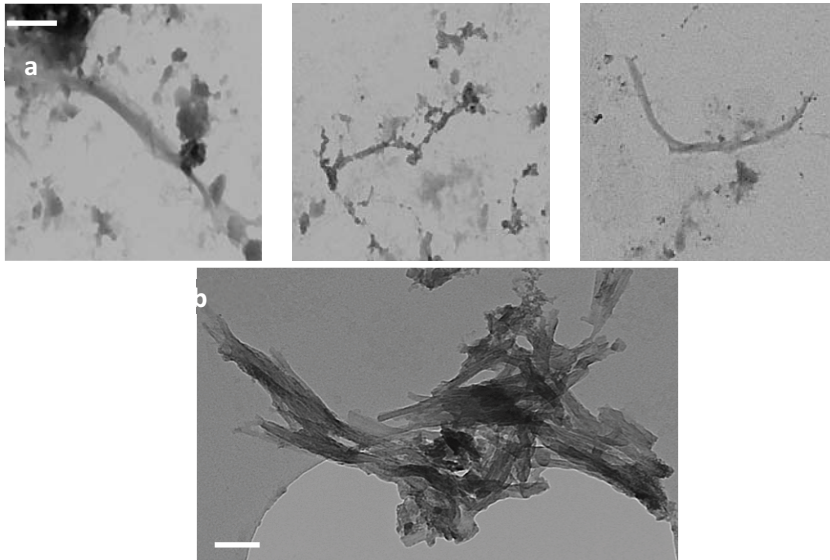


Figure 6-19: EM analysis of captured aggregates from 97Q mouse muscle tissue. Representative AR aggregates in total tissue homogenates (a) or (b) Seprion captured from 97Q muscle tissue homogenate from animals of 17w old were visualized. The captured aggregates in (b) indicated that the sticky fibrillar species in muscle cells come together and this might be a reason for muscle cell function disruption considering the highly specific form and shape of the muscle cells.

We detected the aggregates in Seprion separated material primarily (Fig. 6-19b), but due to very little quantity of the enriched material, we optimized the condition to use total tissue homogenates for our TEM studies, as shown in figure. 6-19a, the extended fibrillar aggregates do present in tissue homogenate and similar to AFM data, they increase in their size and numbers as the animal ages. Our success in optimization of the condition of TEM imaging of *in vivo* material helped us to move one step forward and look into immuno-TEM images of fibrils stained with AR specific antibodies. In figure 6-20, fibrillar aggregates in intact 97Q muscle mouse tissue homogenate look identical to *in vitro* fibrils of amyloid proteins. The immuno-gold dots on the fibrils look to be few but significant. We did all different control conditions to come to this point that the fibrillar species in tissue comprise AR, however as it is visible in TEM images, there are also other proteins involved in formation of fibrillar species. Usually it is challenging to find a condition that antibodies can detect their binding epitopes in amyloid proteins, as more often the binding epitopes enter to the core of aggregates and become inaccessible.

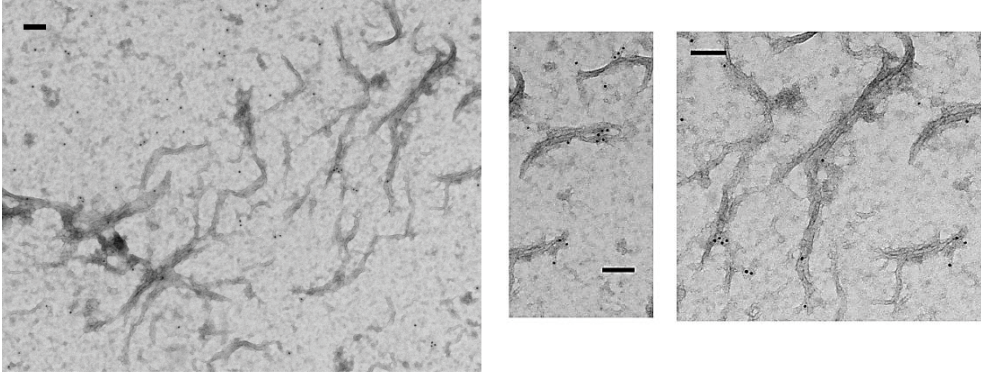
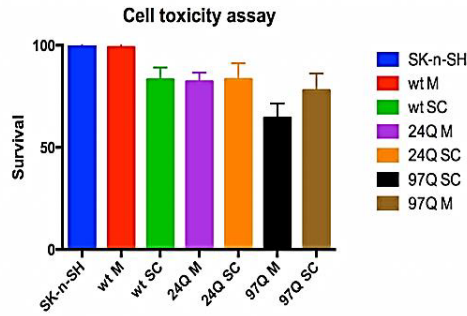


Figure 6-20: Immuno-EM analysis of captured material from 97Q mouse muscle tissue. Representative immuno-gold labeling of AR aggregates with H280 antibody from 97Q muscle tissue homogenate from animals of 14w old. Scale bars are 100nm.

Our Immuno-TEM data in collaboration with Dr. Janet Kumita in Department of Chemistry in University of Cambridge, shows that fibrillar aggregates in muscle tissue of 97Q male SBMA mouse contain AR but in high probability also other co-aggregating proteins. The Immuno-gold labeling of the samples were performed on intact total tissue homogenates to avoid any possible changes to the structure of fibrils. Our visualized data summarize that fibrillar species in muscle tissues contain AR and accompaniment of other proteins, which we assume that might be the reason for the toxic role of these species in muscles.

6.2.8 Only muscle tissue homogenate of SBMA transgenic mice induced cell toxicity in Neuroblastoma cells

To identify whether distinct species in total tissue homogenates of Wt and transgenic animals are toxic for the cells, we seeded 10,000 cells in 96-wells plate and treated them with identical concentrations of total tissue homogenates. Finally we analyzed the activity of the cells after 24h. This type of cell toxicity experiment is only based on apoptosis and therefore does not give any detailed information on the main involvement of specific signaling pathway in toxicity or disrupted cascades, however enough informative for our primary observations. Interestingly, our results illustrated that only 97Q muscle tissue homogenate induce toxicity, whereas the spinal cord tissue did not show any significant difference to Wt and 24Q tissues (Fig. 6-21). This experiment stressed that some variants of AR, either a specific structural variant or truncated variants present in the muscle tissue and not in the spinal cord, which cause cell toxicity in the neuroblastoma.



20%w/v total tissue homogenate	
ANOVA summary	
F	3.802
P value	0.0019
P value summary	**
Are differences among means statistically significant? (P < 0.05)	Yes
R square	0.1904

Dunnett's multiple comparisons test	Mean Diff.	95% CI of diff.	Significant?	Summary	A-?	
SK-n-SH vs. wt M	0.9806	-24.23 to 26.19	No	ns	B	wt M
SK-n-SH vs. wt SC	16.80	-8.412 to 42.02	No	ns	C	wt SC
SK-n-SH vs. 24Q M	17.79	-7.427 to 43.00	No	ns	D	24Q M
SK-n-SH vs. 24Q SC	16.69	-8.523 to 41.90	No	ns	E	24Q SC
SK-n-SH vs. 97Q M	35.53	12.28 to 58.77	Yes	***	F	97Q M
SK-n-SH vs. 97Q SC	22.08	-3.133 to 47.29	No	ns	G	97Q SC

Figure 6-21: AR dependent cytotoxicity of fibrillar species formed in muscle cells and not in spinal cord in SK-n-SH. Neuroblastoma cells were treated with identical concentration of muscle and spinal cord total tissue homogenates of three groups of animals, Wt, 24Q and 97Q. Only 97Q muscle tissue homogenates show significant cytotoxicity in neuroblastoma cells in our experimental condition. Summary of ANOVA and Dunnnett test are indicated.

6.2.9 AR's toxicity in muscle tissue homogenate is due to the formation of intranuclear aggregates

Further investigation was necessary to explain the cause of muscle tissue toxicity. Therefore, 25,000 cells were seeded in Lab Teck™ chambers and were treated with tissue homogenates from Wt, 24Q and 97Q animals. As it is shown in figure 6-22, lab Teck® chambers are very appropriate for immunofluorescent staining. After 24h and 48h, the cells were washed and fixed for immunostaining.

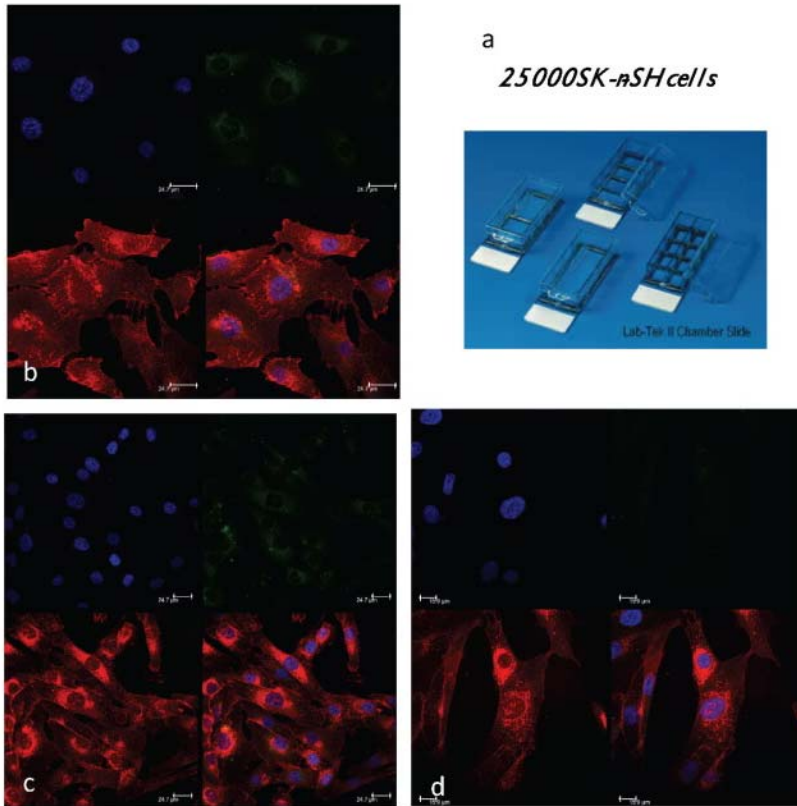


Figure 6-22: a) Representative picture of a Lab TeckTM chamber used for Immuno-Fluorescence staining. b) Immuno-fluorescence images of neuroblastoma cells treated with 24Q muscle (upper left), 24Q spinal cord (lower left) and Wt (right) muscle tissue homogenates. 25.000 cells in Lab Teck chambers were treated with identical concentration of total protein from Wt and 24Q muscle tissues. DAPI nuclear staining in blue, cell membranes staining in red and AR staining in green are indicated.

As it was shown in figure 6-22, tissue homogenates of Wt and 24Q mouse did not show any positive staining in cells, except for very slight positive staining in cytoplasm of 24Q treated cells. It is to note that AR is a nuclear receptor, however in Wt and 24Q, no positive staining was found co-localized with DAPI staining.

In cells treated with 97Q tissue homogenates, both spinal cord and muscles contains some fragments of AR, which can enter the cells (Fig 6-23 and 6-24). Strikingly, only in the cells treated with muscle tissue homogenate, co-localization of DAPI (nuclear) staining and AR positive staining was observed (Fig. 6-24c). We hypothesize that the nuclear positive staining is the indication that aggregated species could be taken up and co-aggregate with endogenous AR. The aggregated species in nucleus could recruited functional nuclear proteins and be the reason for cell toxicity. While in spinal cord tissue, neither nuclear aggregates nor toxicity were observed.

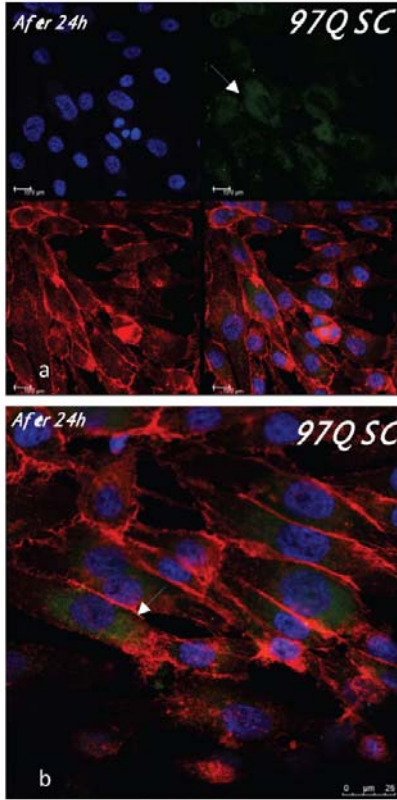


Figure 6-23: a) Representative Immuno-fluorescence staining images of neuroblastoma cells treated with 97Q spinal cord tissue for 24h. b) Higher magnification of neuroblastoma cells treated with spinal cord total tissue homogenate. DAPI nuclear staining in blue, cell membranes staining in red and AR staining in green are indicated. Positive green cytoplasmic staining is indicated in both images.

Further studies will be needed to explain the properties of these aggregates and their ability to enter to cytoplasm, while at the same time no entry to the nucleus occurs. We hypothesize that the observed structural differences by AFM and TEM are the main reason for the observed difference in toxicity behavior.

In the following figure, 6-24, neuroblastoma cells were imaged after 24h and 48h. In muscle tissue treated cells, after 24h, the aggregated species in cytoplasm are not the diffuse observed aggregates of 24Q and 97Q spinal cord tissue, but rather they appear as more granulated and punctuated species with closer proximity to the nucleus. Interestingly, when these cells were incubated for 48h, these aggregated species entered to nucleus and showed co-localization with DAPI staining.

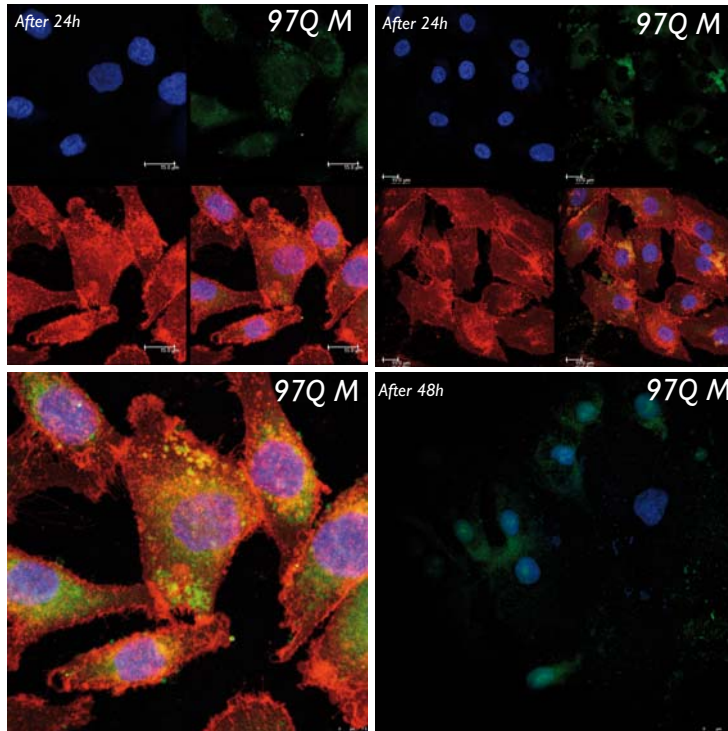


Figure 6-24: AR aggregates enter the nucleus and cause cell toxicity. a) and b) two Immunofluorescence images of two independent experiments where human AR was stained in neuroblastoma cells treated for 24h with muscle total tissue homogenate from 97Q SBMA mice. c) Higher magnification of cells with positive stained punctuated AR aggregates in cytoplasm and nucleus. d) After 48h incubation of cells with muscle tissue homogenates, the aggregates appear to be co-localized with DAPI nuclear staining. DAPI nuclear staining in blue, cell membranes staining in red and AR staining in green are indicated. Positive cytoplasmic staining is indicated in both images.

These results show for the first time that AR with elongated polyQ has cell toxicity properties, not only due to the presence of higher number of Q residues, which makes the protein aggregate, but also due to the specific structural properties that aggregated AR adapts in mouse muscle cells. Moreover, these results illustrate that nuclear aggregates are the toxic species while cytoplasmic AR does not show any toxicity in neuroblastoma cells.

Our study showed that cell specificity does play an important role and this explains that, although AR expresses in all tissues in body (272), aggregation happens only in some specific cells. We believe that this behavior relates directly to proteome of the cells. Cell specificity in neurodegenerative disease is a topic that has been little explored (273). Few studies in well-studied neurodegenerative disease, like Alzheimer's and Parkinson's disease have been done (274–276), while the necessity of these studies is obvious for the field. Therefore, we hope that further studies in this area could provide us with more explanations in the cell toxicity and disease mechanisms in different neurodegenerative diseases.

6.2.10 X-ray phase contrast characterization of muscular degeneration in young and old SBMA transgenic mice

Our step by step approach to illustrate the pathology of SBMA in more details brought us to high resolution X-ray phase contrast micro-tomography method. As briefly described, X-ray phase contrast methods were widely used in the clinic (277–279), but not for high-resolution studies. During the last decade, improvements in this method have made scientists use the concept of tomography to go beyond histology (280, 281). Accurate knowledge of the morphology of the tissue is required for minimally or non-invasive surgical interventions. On the (sub-) cellular level, human tissue is generally characterized using optical microscopy, which allows extracting morphological features with a wide spectrum of staining procedures. The preparation of the histological slices, however, often leads to artifacts resulting in imperfect morphological data. In addition, the generation of 3D data from those is time-consuming. We used synchrotron radiation-based micro computed tomography (SR μ CT) avoiding preparation artifacts and giving rise to the 3D morphology of features such as muscle fiber and neuromuscular junctions on the micrometer level. One can differentiate between appearance, shape and all physical and structural properties of tissue without any staining procedure because of different X-ray absorption values. The tomography data allows quantifying the local strains in the histological images using registration algorithms.

Micro-tomography started to be used as a methodology for biological tissue characterization (282). The advantage of this method over histology is that no samples preparation step is needed, which not only makes it less time consuming, but it also keeps the tissue more intact.

Before data acquisition, muscle and spinal cord tissues were collected from 97Q animals in two age groups of young (young: Pre-onset) animals and animals with no cage activity (old). The tissues were 2-3 mm in diameter and 4-5 mm in length. Tissues were fixed for 24h in 4% PFA and the samples were aligned vertically in agarose gel. This work was done in collaboration with Dr. Ralf Hofmann and Dr. Tobias Jochum in Karlsruhe Institute of Technology (KIT). The data was acquired in synchrotron in Chicago (Advanced Photon Source (APS)) and analyzed in (KIT) in Germany.

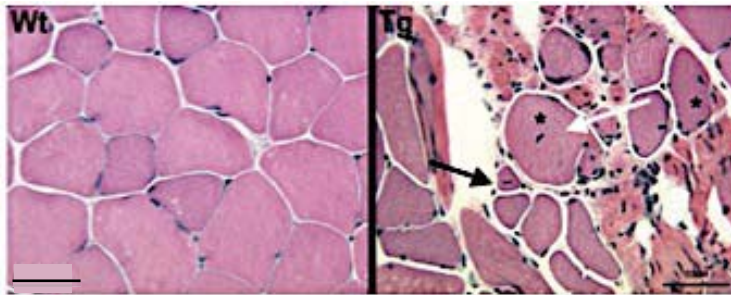


Figure 6-25: Overexpression of Wt AR in muscle fibers of transgenic male mice leads to muscle pathology. H&E staining of Extensor Digitorum Longus (EDL) muscle of Wt and transgenic male mice indicate that transgenic muscles contain both abnormally small, angular fibers (black arrow) and hypertrophied fibers (white arrow). Both fiber types contain internal nuclei (stars) Adapted from Michigan State University webpage, The Jordan lab).

In figure 6-25, representative hematoxylin and eosin (H&E) staining images of (On left) healthy muscle tissue with expression of Wt AR and (on right) an AR overexpression mouse muscle tissue were compared. In healthy skeletal muscle tissue, the cells maintain a well round shape with a very thin layer of connective tissue that separates the cells, while on the right side, the cells shrink to abnormally small cells with angular forms and hypertrophied fibers with replacement of thick layer of connective tissue (10, 283, 284).

Our observation from the conventional histology methods and our interest in more high-resolution data, we investigated more the muscle tissue properties with X-ray.

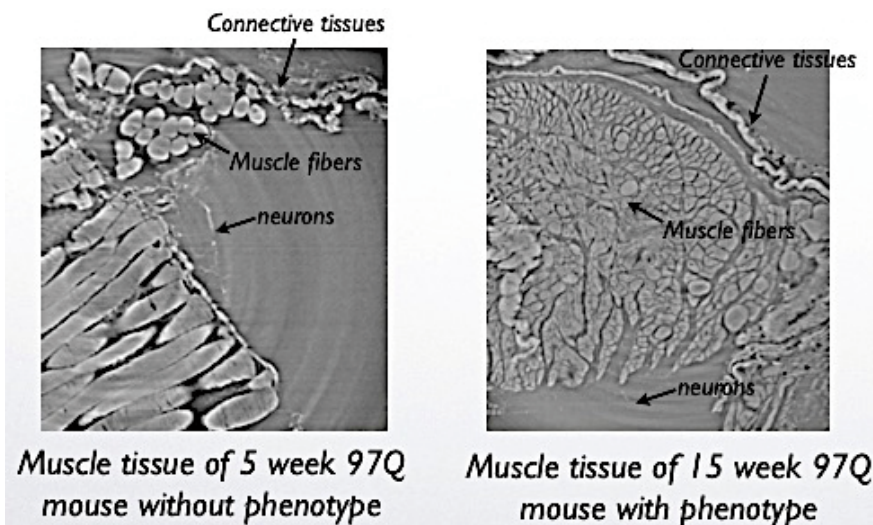


Figure 6-26: Visualization of muscle tissue properties in 5w and 15w old 97Q mice. A representative stack of sagittal cut of muscle tissues from young (left) and old (right) 97Q mice indicate that the healthy muscle fibers (round in sagittal and extended in longitudinal cut) decrease in number in old mouse with no cage activity. In old muscle tissue only very few numbers of round, well shaped fibers are visible (black arrow). The connections of neurons also indicated by black arrows.

In figure 6-26, acquired and analyzed data by synchrotron from two stacks of a young and an old mouse muscle tissue are shown. The image on the left side visualizes 5w old mouse (pre-symptomatic) with healthy muscle tissue features (285). On the right side, in old 97Q mouse a clear degeneration of cells is visible (286). The synchrotron data contains over 10000 stacks that were acquired through the whole length of tissue. This implies that each stack contains information that can be followed sequentially in other stacks. Using this approach, we followed the nerve branching systems in specific selected regions of interest (Neuromuscular Junction) in 10000 stacks and every 50 stacks we reconstructed the image (280, 281). Finally, all the reconstructed images were merged in one single image (Fig. 6-27).

3D reconstruction of skeletal muscle of 97Q young mouse

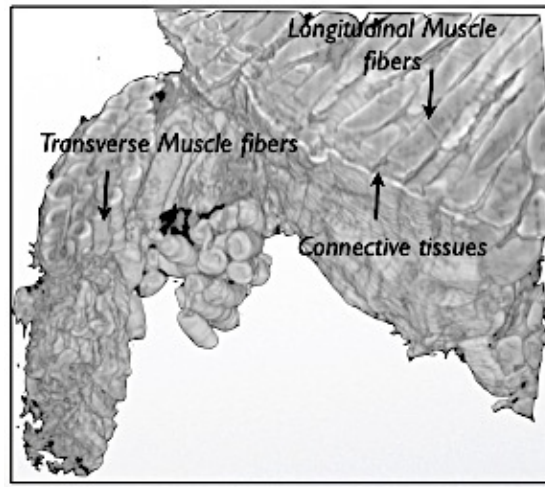


Figure 6-27: 3D reconstructed image from sum of series of axial stacks of a young 97Q mouse muscle tissue. Two different views of transverse muscle fibers and longitudinal muscle fibers were reconstructed from series of images acquire from a 5w old mouse muscle tissue.

As it was shown in figure 6-27, the acquired data from synchrotron provides us with high-resolution information on muscle cells properties. As illustrated, in young 97Q mouse muscle tissue, the healthy muscle fibers usually look either extended in longitudinal cut or well-shaped round in transverse plane (287, 288). As well, we followed a small population of nerves through 10000 stacks and summed the images in Figure 6-28. The reconstructed image is presented with colors representative of each tissue type.

Neuromuscular junction (NMJ) is the spot where motor neurons meet the muscle cells. Changes in NMJ are correlated to aging and neuromuscular diseases (66,

289–291). To date, there is no clear data available on the NMJ changes in SBMA pathology, if they change their structures, or increase in number or even disappear or lose their functionalities (292). We were able to show that using this high-resolution method, it is possible to reconstruct and characterize individual NMJ, not only with regard to appearance of NMJ but also in a quantitative way. Further characterizations of muscles and spinal cord tissues are planned for future.

3D reconstruction of neuromuscular junction on skeletal muscle of 97Q young mouse

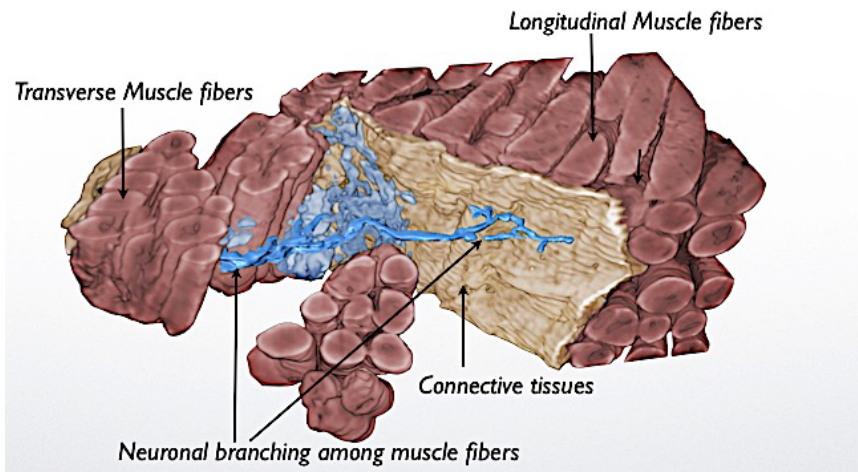


Figure 6-28: 3D reconstruction of neuronal branching and neuromuscular junction in a 97Q young muscle tissue. Connective tissue in cream, muscle tissues in reddish brown and nerves in blue were indicated. The image is a reconstruction of a series of axial stacks where we followed a specific nerve region showed in blue here.

Our data shows the possibility of getting high-resolution information from structural changes in muscles as the animal ages. This data will include changes in shapes and forms of muscle cells, differences in number of NMJ and variations in connective tissue properties (293). We hypothesize that by comparing the 5w old 97Q mouse with 15w old mouse of the same genetic background and genotype, we can see details of changes, which correlate with the disease manifestations and pathology.

6.2.11 High-resolution single molecule microscopy, a better way to illustrate the fibril structures

The details of STORM and STED microscopy have been previously discussed in M&M. Our previous data from TEM and AFM provided us with a detailed characterization of mouse tissue homogenates of 97Q mice. However, our interest

in applying new high-resolution methods, as well as their complementary information motivated us to investigate the fibrillar species in total tissue homogenate using AR specific antibody. This work was done in collaboration with Dr. Pierre Mahou, in the laboratory of Prof. Kaminski at University of Cambridge.

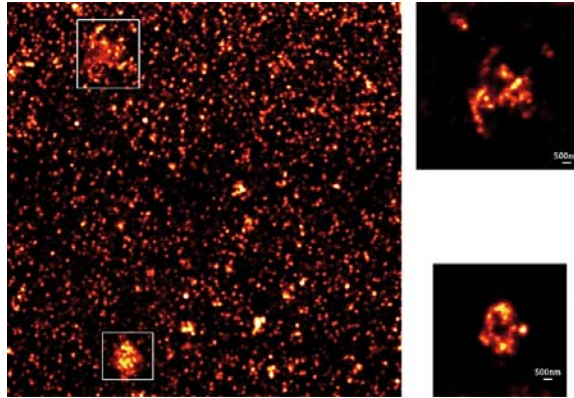


Figure 6-29: PolyQ-mouse-97Q Confocal-Excitation 640nm Power = 100%, FOV = 40 μ m, Pixel size = 20nm

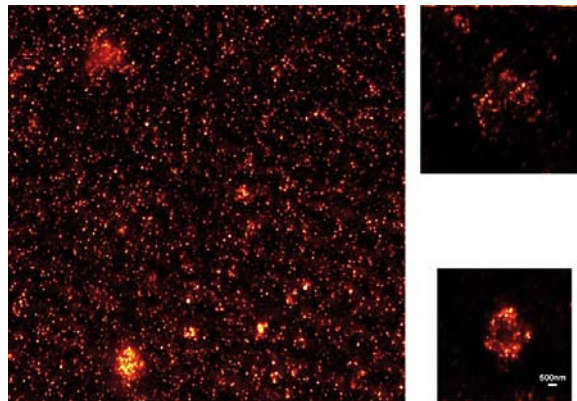


Figure 6-30: PolyQ-mouse-97Q Confocal-Excitation 640nm Power = 100%, Depletion 765nm Power= 100%, FOV = 40 μ m, Pixel size = 20nm, 30 μ s/pixel

Using a specific primary antibody for elongated version of polyQ tract (1C2) and a STED compatible secondary antibody (ATTO 647 anti-mouse antibody), the aggregates in mouse muscle total tissue homogenate were stained and imaged using a 647nm Laser. Aggregated species were intact and the structures were not affected by any pre-treatment or sample preparation procedure. As it is shown in figure 6-29 and 6-30, polyQ positive AR aggregates in muscle tissue form 3D structures of fibrillar species that were only found in muscle tissues. The upper image (Fig. 6-29) is a confocal image of two representative aggregates composed of fibrillar species of polyQ AR, while the lower image is the STED image with improved resolution (Fig. 6-30). The single molecule resolution of STED images

visualizes the arrangements of single positive stained spot along the length of fibrils.

Beside STED microscopy, we looked also into the very similar tissue samples by dSTORM, which in practice is based on blinking of each single molecule, which contains one fluorophore induced by a blinking solution (previously explained in M&M). We imaged 10,000 stacks of each area, while each single point was blinking. The number of the photons received by camera were counted and we used a MATLAB based program to analyze the acquired blinking of 10,000 stacks and to obtain a sum of blinking points (imagining a fibril composed of 100 monomeric proteins that have 100 secondary antibodies blinking on it).

In this method, we stained the aggregates by anti AR specific primary antibody and used an Alexa-fluor568 secondary antibody to visualize AR by fluorescence microscopy. In a conventional way we cannot resolve these 100 points in very close proximity, as they will all blink as one single large species, whereas if we make them blink in different time order, in each image only some of the dyes will blink and eventually all the images will be summed up to provide the last STORM images with resolved information from each individual point.

STORM images have a resolution of 20nm. This means that only the positive stained points in closer proximity than 20nm cannot be high-resolution resolved. In figure 6-31, we compared the muscle and spinal cord total tissue homogenate aggregates in 97Q and 24Q transgenic mice. It is not clear if the AR aggregates are composed of only N-terminal truncated AR or the full length AR. We stained the aggregates with two primary antibodies, one specific for N-terminal and one specific for C-terminal domains. Two secondary antibodies with fluorophore were used. The tissues were directly spotted on lab Tek™ wells and imaged with green laser (568nm) and red (647nm) in presence of specific secondary antibodies.

As it is shown in figure 6-31, in samples from 24Q mice, we have significantly less population of green and red staining compared to 97Q animals. This is another validation of the presence of a higher number of aggregated species in 97Q animals (a and b with c and d).

Figure 6-31 in lower panel illustrates the mean area of the stained AR species by either green (for N-terminal) or red (for C-terminal) in different tissues of 97Q and 24Q. In the spinal cord samples of 97Q and muscle samples of 24Q, no significant difference between the areas of green and red staining were observed. This support our hypothesis that most of the AR presents as full length, whereas in muscle tissue of 97Q animals, the stained area with green antibody specific for N-terminal domain were significantly higher compare to the red stained area, which suggest that most of the aggregates in muscle tissue comprise only the N-terminal truncated variants of AR.

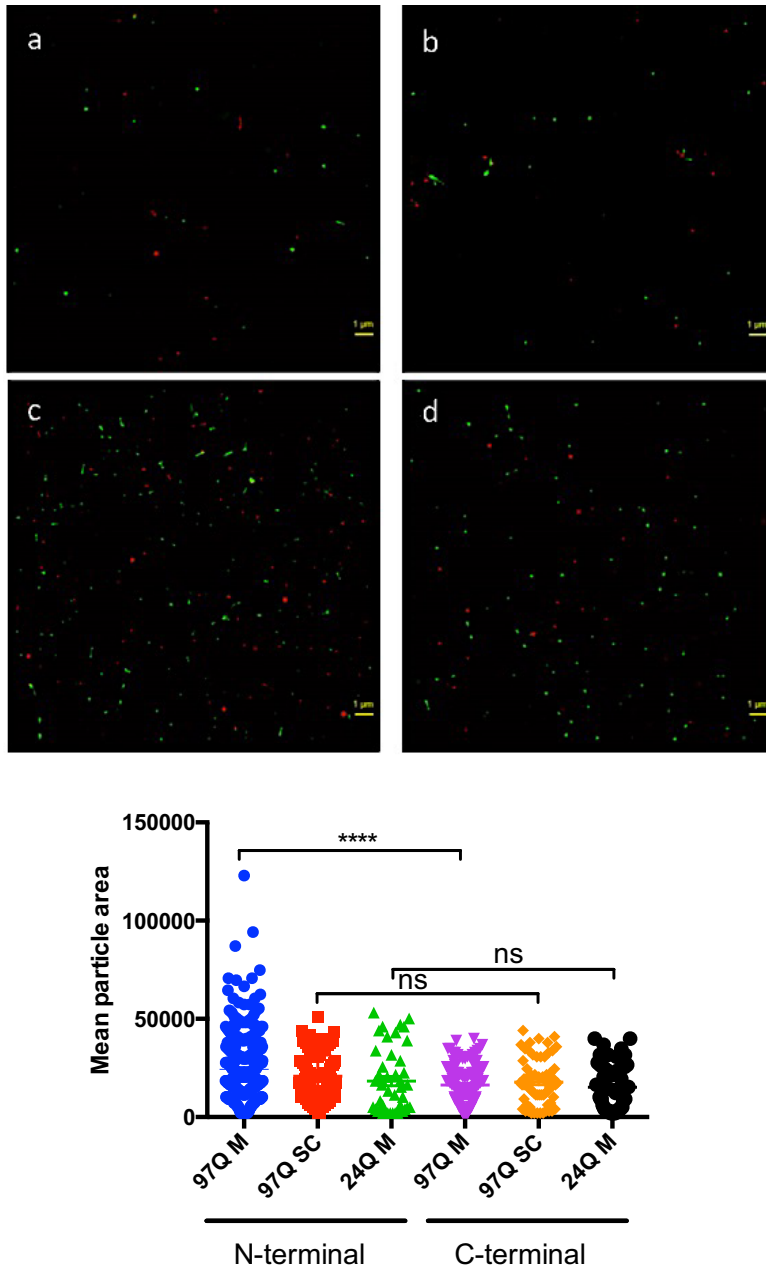


Figure 6-31: (Upper panel) 10%w/v muscle and spinal cord tissue homogenates from two transgenic mouse models of SBMA (either with 24Q (a: muscle and b: spinal cord) or 97Q (c: muscle and d: spinal cord)) were stained with two primary antibodies, mouse anti human AR antibody specific for N-terminal domain and rabbit anti human AR antibody specific for C-terminal. Mouse and rabbit specific secondary antibodies with Alexa 647 and Alexa 568 were used for the dSTORM microscopy. The scale bar is 1 μm in all the images. (Lower panel) mean particle area of AR aggregates in muscle and spinal cord tissues stained with H280 and C19 were compared.

6.2.12 Anle138b: Could a novel oligomer modulator for disease-modifying therapy of neurodegenerative diseases such as prion and Parkinson's disease also affect SBMA mice pathology?

In neurodegenerative diseases such as Alzheimer's disease, Parkinson's disease and prion diseases, deposits of aggregated disease-specific proteins are found. Oligomeric aggregates are presumed to be the key neurotoxic agent (294, 295). Here, we describe the novel oligomer modulator anle138b [3-(1,3-benzodioxol-5-yl)-5-(3-bromophenyl)-1H-pyrazole], an aggregation inhibitor, developed based on a systematic high throughput screening campaign combined with medicinal chemistry optimization at the Max Plank Institute in Goettingen.

In vitro, anle138b blocked the formation of pathological aggregates of prion protein (PrP^{Sc}) (296), of α -synuclein (α -syn) deposited (297) in PD and other synucleinopathies such as dementia with Lewy bodies (DLB) and multiple system atrophy (MSA) (294). Notably, anle138b strongly inhibited all prion strains tested including BSE-derived and human prions. Anle138b showed structure-dependent binding to pathological aggregates and strongly inhibited formation of pathological oligomers *in vitro* and *in vivo* both for prion protein and α -synuclein. Both in mouse models of prion disease and in three different PD mouse models, anle138b strongly inhibited oligomer accumulation, neuronal degeneration, and disease progression *in vivo*. Anle138b had no detectable toxicity at therapeutic doses and an excellent oral bioavailability and blood–brain-barrier penetration. These findings indicated that oligomer modulators provide a new approach for disease-modifying therapy in these diseases, for which only symptomatic treatment is available so far. Moreover, these findings suggested that pathological oligomers in neurodegenerative diseases share structural features, although the main protein component is disease-specific. This indicates that compounds such as anle138b that modulate oligomer formation by targeting structure-dependent epitopes can have a broad spectrum of activity in the treatment of different protein aggregation diseases.

Based on the published data on this compound and its effect on the improvement of disease phenotype in Prion disease and PD transgenic animals, we decided to study the effects of this general oligomer modulator in SBMA transgenic animals. We started with two groups of transgenic animals and one group of Wt mouse as a control (Fig. 6-32). The group of Wt animals didn't show any toxicity as was expected. Survival plots indicated a significant difference between the groups treated with compound and the placebo treated one (Fig. 6-33).

Anle138b: a novel oligomer modulator for disease-modifying therapy of neurodegenerative diseases such as prion and Parkinson's disease

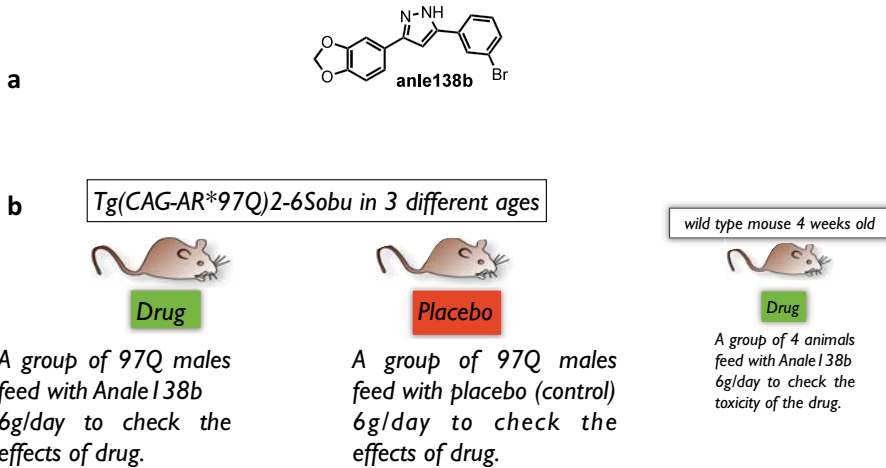


Figure 6-32: a) Chemical structure of anle138b. This compound is chosen from a library of 10,000 chemically diverse drug-like compounds with regard to inhibition of prion protein aggregation in a molecular screening assay. b) Experiment design for life span analysis studies. Two groups of 97Q male transgenic animals with similar age and randomly chosen from different breedings were fed with either placebo (6g/day) or drug (6g/day). One group of Wt mice was just tested for toxicity.

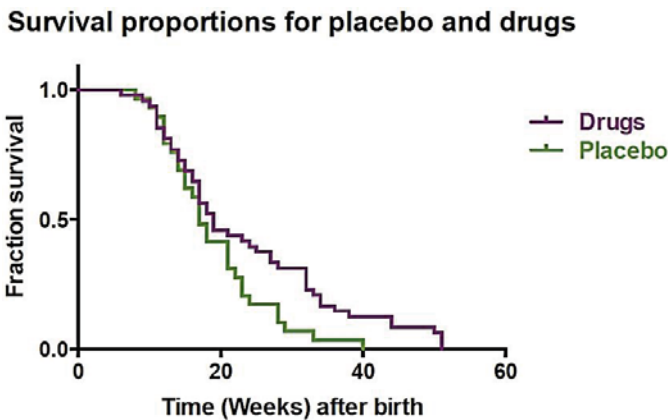


Figure 6-33: Influence of daily anle138b treatment on SBMA pathology transgenic animals. Our life span analysis data did not show a significant difference between the drug (violet) and placebo (green) in animals in the early stage of the disease, while in older animals, it showed a significant (with one *) effect on increasing of the life span. Mantel-Cox test used to compare the survival distribution of two unpaired samples.

Log-rank (Mantel-Cox) test: to compare survival distribution of two unpaired samples Chi square= 4.179, df= 1, P value = 0.0409- P value summary *, Are the survival curves sig different? Yes

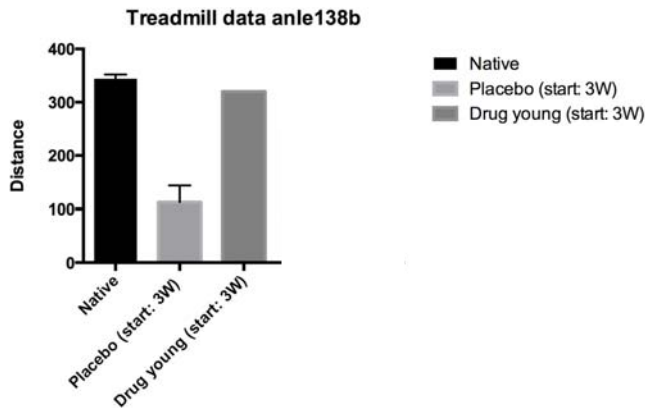


Figure 6-34. Maximal treadmill running distance for mice of indicated groups. Data represented mean \pm s.d of 6 mice per group (20 w old). Three groups of Native (Wt mice), Placebo and drugs all were fed by age of three weeks are indicated. There is a significant difference between placebo and drug treated groups.

In figure 6-33, two groups of 97Q animals (6 animals in each) were fed either by placebo or drug starting at the age of three, six or ten weeks. We followed specific criteria to make the groups, based on their family background. Animals with the same age, no sibling and in healthy conditions at age of three weeks were chosen for experiments. A group of Wt animals with the same age were fed by drug and life span was checked for toxicity. No toxicity was observed in this group and they had a life span similar to non-treated animals.

Life span analysis data from a large number of animals in both group of placebo and drug did not show any significant difference in animals with age less than 20 weeks while it started to show a significant difference as the animals got older (Fig. 6-33).

Two groups of six 97Q mice were trained for three times every other day and on the 4th day were tested. From all the parameters we obtain from treadmill analysis (1- distance, 2-speed and 3- time), we focused on distance. There are some suggested protocols for small animal treadmills, which we optimized for our study. In groups of animals with treatment started from third week after birth, motor function analysis was done using a treadmill test at age of 20 weeks.

Treadmill parameter adjustment:

0.1mA for the electric shock zone

Leave the animals for 5-10 min in the treadmill room

Animals trained for 10 min in 14 cm/s

Each training was in a two days distance

At least six animals in each group

The motor function analysis by treadmill showed that the animals treated by drug have a significantly better performance compared to the animals treated with placebo (Fig. 6-34). It is noteworthy that the treatment showed its effect in animals that got the treatment from the 3rd week after birth, while the drug showed less effect in groups starting with older age (9-10 weeks old- data not shown). Therefore, we hypothesized that in SBMA drug development strategies, we might need to consider that early-stage treatment of the patients is more effective and therefore, it is highly suggested to look for biomarkers of the disease for pre-symptomatic diagnosis of all polyQ diseases, as this could bring us to earlier and more effective therapeutic strategies.

6.2.13 AR and its interacting partners

IDPs, while structurally poor, are functionally rich by virtue of their flexibility and modularity (120, 298). However, how mutations in IDPs elicit diseases, remain to discuss. For AR function as a transcription factor; it is necessary to activate the whole transcription machinery system with its almost 500 residue intrinsically disordered domain. Using published data in literature; we elucidated the binding partners of AR N-terminal domain that might play a role in SBMA disease mechanisms (299, 300). We started with a simple approach and used bioinformatics tools to look into all interacting proteins already suggested in literature. We generated a primary interactome of AR and its 318 interacting protein shown partly in figure 6-36 and 6-37. This data was transferable to other bioinformatics tools to answer more biologically relevant questions. Using bioinformatics plugin (301, 302), we analyzed the data and distributed the information in groups based on their molecular functions; cellular compartments they are involved in and biological processes they are involved in (Fig. 6-37). The idea was to take the literature-extracted list and add experimental data to it (Fig. 6-38).

Therefore, we used Cytoscape program (UCSD, California) to integrate biomolecular interaction networks of AR from literature with high-throughput expression data and other molecular states, into a unified conceptual framework. Although applicable to any system of molecular components and interactions, Cytoscape is most powerful when used in conjunction with large databases of protein-protein, protein-DNA and genetic interactions that are increasingly available for humans and model organisms (302–304). We prepared a list of AR-interacting protein from literature, and made a network of protein-protein interaction (Fig. 6-35 and 6-36). Cytoscape gives a visualization of interactions and the components of nodes (interacting proteins) and edges (interactions), which can be valued based on their probability of being true. In figure 6-37, an example of AR interacting protein and a zoom in the main region with visualized nodes and edges, are indicated. It is to note that nodes and edges in Cytoscape are not only connected to AR, but also contain information about the interacting partners. The

whole network prepared in Cytoscape was analyzed also in a Plug in of Cytoscape, named BiNGO. The Biological Networks Gene Ontology tool (BiNGO) is a Java tool to determine which Gene Ontology (GO) terms is significantly overrepresented in a set of genes. BiNGO can be used either on a list of genes, pasted as text, or interactively on subgraphs of biological networks visualized in Cytoscape. BiNGO maps the predominant functional themes of the tested gene set on the GO hierarchy (301), and takes advantage of Cytoscape's versatile visualization environment to produce an intuitive and customizable visual representation of the results (302, 304). Figure 6-38 and 6-39 are two examples of the presentation of data in BiNGO. It gives a quick overview in which cascades and biological component, AR and its interacting partners are over presented. The primary interactome of Wt AR was used to extract interacting protein data for further analysis of experimental data.

But we realized that the bioinformatics data from our initial interactome need to be optimized for specific proteins which interact with NTD or are involved in aggregation formation of AR. Our hypothesis, as it was pointed out in this work, similar to some other groups, is that AR sequester many other critical and functional proteins while it is aggregating (239, 305). Based on our NMR data and discussed coiled-coil interactions (47), we believe that the sequestration of other proteins into polyQ aggregates is very selective with importance for both structure and sequence of these proteins (45).

ID	Column 1	Column 2	Column 3	canonicalName	cytoscape.alias.list	NTD interacto
NSD1	NSD1	Q96L73	Histone-lysine N-methyltransferase, H3 lysine-...	NSD1	[NSD1]	yes
AES	AES	Q08117	Amino-terminal enhancer of split	AES	[AES]	yes
GTF2F2	GTF2F2	P13984	General transcription factor IIF subunit 2	GTF2F2	[GTF2F2]	yes
DDC	DDC	P20711	Aromatic-L-amino-acid decarboxylase	DDC	[DDC]	yes
CDK7	CDK7	P50613	Cyclin-dependent kinase 7	CDK7	[CDK7]	yes
CCNE1	CCNE1	P24864	G1/S-specific cyclin-E1	CCNE1	[CCNE1]	yes
SFPQ	SFPQ	P23246	Splicing factor, proline- and glutamine-rich	SFPQ	[SFPQ]	yes
STAT3	STAT3	P40763	Signal transducer and activator of transcription 3	STAT3	[STAT3]	yes
COX5B	COX5B	P10606	Cytochrome c oxidase subunit 5B, mitochondrial	COX5B	[COX5B]	yes
PIAS3	PIAS3	Q9Y6X2	E3 SUMO-protein ligase PIAS3	PIAS3	[PIAS3]	yes
HSPB2	HSPB2	Q16082	Heat shock protein beta-2	HSPB2	[HSPB2]	yes
NRIP1	NRIP1	P48552	Nuclear receptor-interacting protein 1	NRIP1	[NRIP1]	yes
CASP8	CASP8	Q14790	Caspase-8	CASP8	[CASP8]	yes
IDE	IDE	P14735	Insulin-degrading enzyme	IDE	[IDE]	yes
CREBBP	CREBBP	Q92793	CREB-binding protein	CREBBP	[CREBBP]	yes
SVIL	SVIL	Q95425	Supervillin	SVIL	[SVIL]	yes
NCOA4	NCOA4	Q13772	Nuclear receptor coactivator 4	NCOA4	[NCOA4]	yes
NR2C2	NR2C2	P49116	Nuclear receptor subfamily 2 group C member 2	NR2C2	[NR2C2]	yes
GSK3B	GSK3B	P49841	Glycogen synthase kinase-3 beta	GSK3B	[GSK3B]	yes
NCOA2	NCOA2	Q15596	Nuclear receptor coactivator 2	NCOA2	[NCOA2]	yes
RB1	RB1	P06400	Retinoblastoma-associated protein	RB1	[RB1]	yes
PKN1	PKN1	Q16512	Serine/threonine-protein kinase N1	PKN1	[PKN1]	yes
NCOR1	NCOR1	Q75376	Nuclear receptor corepressor 1	NCOR1	[NCOR1]	yes
PA2G4	PA2G4	Q9UQ80	Proliferation-associated protein 2G4	PA2G4	[PA2G4]	yes
APPBP2	APPBP2	Q92624	Amyloid protein-binding protein 2	APPBP2	[APPBP2]	yes
GTF2F1	GTF2F1	P35269	General transcription factor IIF subunit 1	GTF2F1	[GTF2F1]	yes
CEBPA	CEBPA	P49715	CCAAT/enhancer-binding protein alpha	CEBPA	[CEBPA]	yes
MACROD1	MACROD1	Q98069	MACRO domain-containing protein 1	MACROD1	[MACROD1]	yes
RBM14	RBM14	Q96PK6	RNA-binding protein 14	RBM14	[RBM14]	yes
CALM1	CALM1	P62158	Calmodulin	CALM1	[CALM1]	yes
KDM1A	KDM1A	O60341	Lysine-specific histone demethylase 1A	KDM1A	[KDM1A]	yes
PQB1	PQB1	O60828	Polyglutamine-binding protein 1	PQB1	[PQB1]	yes
CDK11B	CDK11B	P21127	Cyclin-dependent kinase 11B	CDK11B	[CDK11B]	yes
NCOR2	NCOR2	Q9Y618	Nuclear receptor corepressor 2	NCOR2	[NCOR2]	yes
NR0B2	NR0B2	Q15466	Nuclear receptor subfamily 0 group B member 2	NR0B2	[NR0B2]	yes
ZNF318	ZNF318	Q5VUA4	Zinc finger protein 318	ZNF318	[ZNF318]	yes
TNK2	TNK2	Q07912	Activated CDC42 kinase 1	TNK2	[TNK2]	yes
RANBP9	RANBP9	Q96559	Ran-binding protein 9	RANBP9	[RANBP9]	yes
NONO	NONO	Q15233	Non-POU domain-containing octamer-binding...	NONO	[NONO]	yes
ETV5	ETV5	P41161	ETS translocation variant 5	ETV5	[ETV5]	yes
CAV1	CAV1	Q03135	Caveolin-1	CAV1	[CAV1]	yes
FOXH1	FOXH1	Q75593	Forkhead box protein H1	FOXH1	[FOXH1]	yes
NCOA3	NCOA3	Q9Y6Q9	Nuclear receptor coactivator 3	NCOA3	[NCOA3]	yes
GAK	GAK	O14976	Cyclin-G-associated kinase	GAK	[GAK]	yes
ZMIZ1	ZMIZ1	Q9ULJ6	Zinc finger MIZ domain-containing protein 1	ZMIZ1	[ZMIZ1]	yes
MAGEA11	MAGEA11	P43364	Melanoma-associated antigen 11	MAGEA11	[MAGEA11]	yes
MCM7	MCM7	P33993	DNA replication licensing factor MCM7	MCM7	[MCM7]	yes
ESR1	ESR1	P03372	Estrogen receptor	ESR1	[ESR1]	yes
DAXX	DAXX	Q9UBP7	Death domain-associated protein 6	DAXX	[DAXX]	yes

Figure 6-35: An example of Data panel window from Cytoscape program. All the data extracted from literature on AR interacting proteins were loaded and a network was built in Cytoscape. This will give an overview of all interacting proteins based on chosen criteria (in this case interaction with NTD of AR).

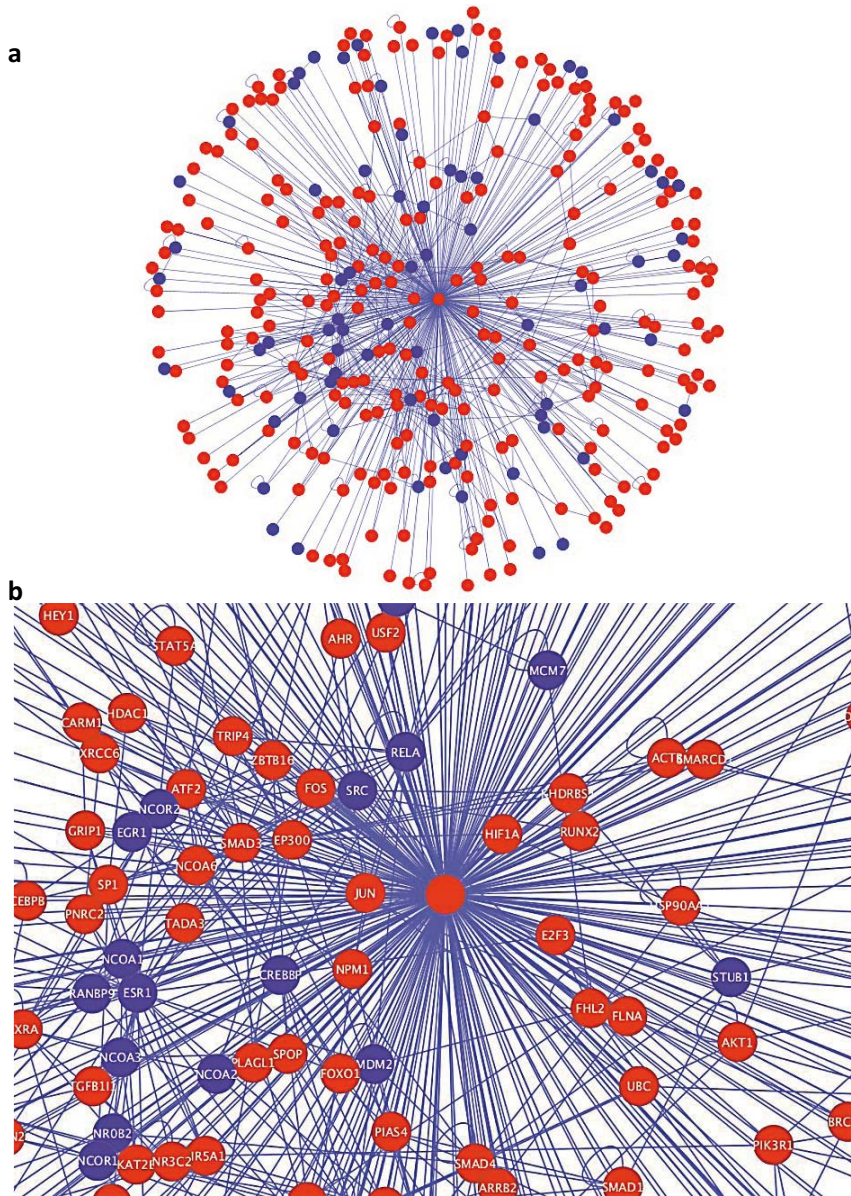


Figure 6-36: Wt AR N-terminal domain primary interactome visualized by Cytoscape. a) 318 interacting protein with N-terminal domain of AR has been identified in literature. Based on our proteomics analysis, there are more proteins that are interacting with N-terminal domain which are yet to be determined. b) a zoom of the central part of interactome. The proteins with direct interaction with AR-NTD are identified in red circle (nodes) and the ones with indirect interaction are shown in violet. (Visualization tool: Cytoscape).

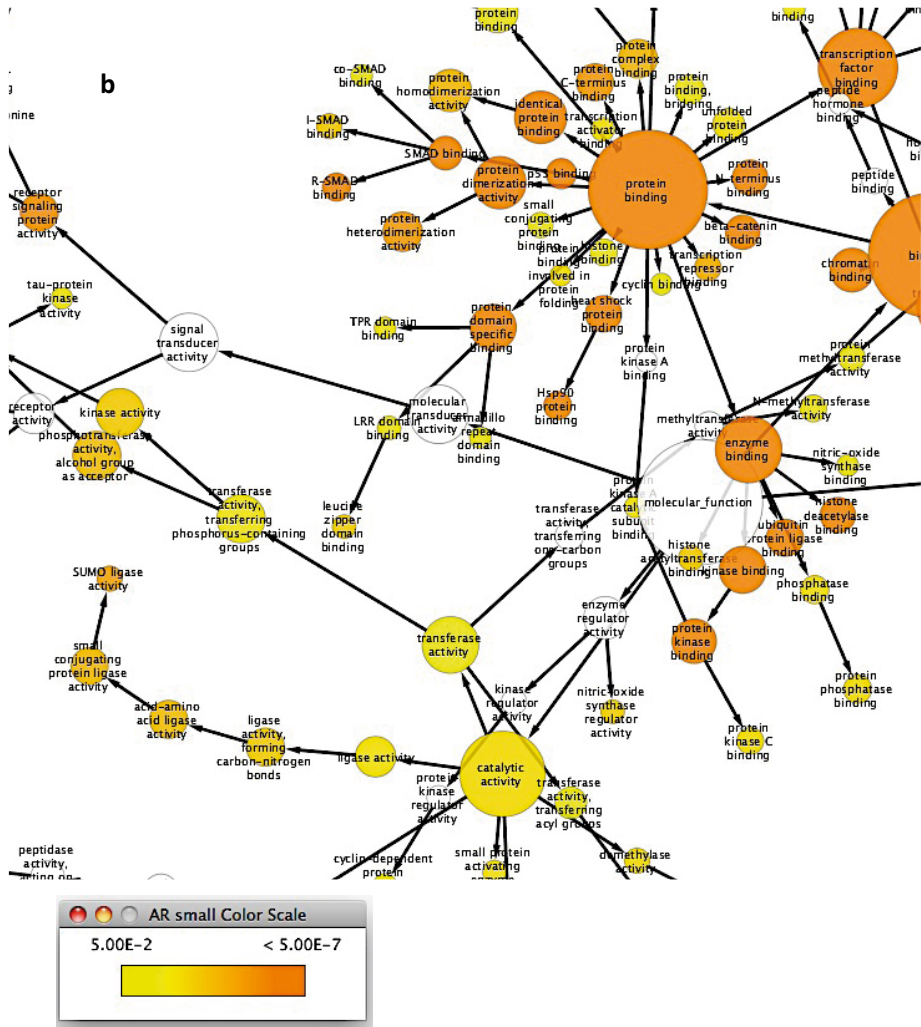


Figure 6-38: Overrepresented gene ontology categories in the set of AR interactome. a) Complete overview of BiNGO data from AR interactome. b) Zoom in a representative region. BiNGO (the Biological Network Gene Ontology tool) analyses representing over-represented categories of the ontology Biological Process among the genes that are marked by AR and significantly interact with AR. The size of the nodes is proportional to the number of genes in the test set which are annotated to that node. Colored nodes are significantly over-represented, with a color scale ranging from yellow (p -value = 0.05) to dark orange (p -value = $5.00E-7$). Statistical testing was as described by Maere et al. (2005)(301).

6.2.14 PolyQ expanded protein and its interacting partners

Intriguingly, although proteins mutated in polyQ diseases have very different biological functions and are widely expressed in various tissues, their mutant forms cause neurodegenerative diseases, preferentially in specific muscular and neuronal cell types. A major question that remains unanswered is why particular cells are more susceptible to degeneration than others (some ideas were previously discussed). For example, in the case of SCA1 from the polyQ diseases family, which is caused by mutant ATXN1 protein a gradual atrophy and loss of cerebellar Purkinje neurons is observed (306). PolyQ-expanded ATXN1 likely destabilizes its interaction with the partner protein ROR- α (307), which is a known regulator of development and function of Purkinje neurons (308).

As previously discussed at end of chapter 4, our observations suggest that abnormal interactions between a polyQ-expanded protein and its interacting partners might be responsible for toxicity and cell loss in polyQ diseases (309). A characteristic pathological hallmark in polyQ diseases is the formation of large inclusion bodies with insoluble protein aggregates in the nuclei of affected neurons. These inclusions mostly consist of polyQ proteins, but also contain ubiquitin and components of the protein clearance machinery (310).

The idea of the polyQ expanded protein and its interacting partners, originally proposed by Perutz (311, 312) and supported by data from the labs of Housman, Paulson, and others, stated that the mutant polyQ assumes an altered conformation and binds to short polyQ stretches normally present in many transcription factors and transcriptional co-activators, such as Cyclic Adenosine Monophosphate Response Element Binding Protein (CBP) (61). This polyQ interaction is postulated to alter CBP's structure. Cell transfection studies have shown that this effect can remove CBP from its normal site of survival-promoting action within cells, causing toxicity (61, 64, 121). This is only seen with the mutant protein, consistent with the idea of a gain of a novel function. However, at present, *in vivo* support for this mechanism is still missing. When considering the issue of aggregation, it may be too simplistic to limit the models for the postulated toxic species to either soluble monomers or insoluble fibrillar aggregates. It suggests that not only the structural aspects of the protein (aggregated or monomeric) are important, but also the protein contents and their interacting partners. The observation that Wt-AR can also aggregate and re-capitulate the SBMA manifestations (84) shows that cell toxicity might be mainly due to its interacting partners due to aggregation. However, the interacting partners -the ones that will be captured to aggregates- are not critical for cellular function, the cells can cope with them.

6.2.15 Identification of AR aggregates associated proteins

The neurotoxicity of expanded polyQ proteins involves structural transitions, mostly believed, from α -helical structure to β -sheets. As it has been previously explained, it is known that polyQ proteins sequester other Q/N rich proteins (305), but no experimental data yet supported this. It was also emphasized before that coiled-coil interaction has one of the main decisive roles in selecting the interacting partners of polyQ proteins (45). Previous work on the analysis of interacting partners of polyQ proteins has shown a recurrent presence of coiled-coil domains in segments that overlap or flank the polyQ region. Our studies on characterization of tissue homogenates of SBMA transgenic mice as well, showed that AR is not the only species present in aggregates, and that other proteins co-aggregate with AR, which could cause the toxicity.

For all polyQ diseases, the protein-induced toxicity is believed to be due to aberrant protein interactions, followed by loss of function of cellular cascades. Accordingly, identification of AR aggregates-associated proteins can provide a powerful tool to assess the involvement of elongated polyQ AR in different cellular pathways. To date, due to complexities in proteomics analysis of aggregation prone proteins, mainly due to their intrinsic properties in being resistant to enzymatic digestions, no studies have identified meaningful AR interactome, neither for polyQ-AR nor for Wt AR involved in prostate cancer.

There are some studies with other polyQ proteins like Htt (313), but none can be directly adapted to our tissues of interest, muscle and spinal cord. Therefore, we optimized all the conditions to identify the proteins that are involved in aggregates of AR in muscles and spinal cord. As it has been previously also emphasized, we were interested in the time evolution studies of how a process evolves along the time. Hence, the main aim of this study was to see if we find a trend in proteins that co-aggregate or co-present in polyQ aggregates analyzed by mass spectrometric approach.

In this case, we were interested in identifying the proteins that show some elevation in the SBMA 97Q mouse tissue homogenates, which includes either a significant decrease or an increase in abundance of some proteins, as the mice age. We focused on a significant number of 97Q mice muscle tissue homogenates from three age groups of pre-symptomatic mice with age around 4-6 weeks, middle age mice and the ones with no cage activity around 15-17 weeks old. The total tissue homogenates from muscle and spinal cord were prepared as explained before and were analyzed By Ricard Rodriguez-Mias in the proteomics facility at the Washington University in Seattle.

Data analyses from three groups of young, middle age and old 97Q muscle tissue homogenates showed striking results. The data demonstrated that comparing these groups, there are some proteins that have an extensive decrease or increase as the mice age. From the preliminary list of almost 2000 proteins found in

aggregates mixture of AR (in index), we identified around 250 proteins with either extremely lower abundance or higher abundance in tissue homogenates. Detailed data analysis of the list demonstrated that most of the proteins, which become extremely affected by aging (either increase or decrease) are involved in physiological muscle functions (Fig. 6-39), either as structural proteins of muscle (for muscle contraction) or proteins involved in muscle metabolic functions.

These data may serve as a further validation for the importance of muscle cells and their very specific structure and function in SBMA. Ultimately, we can hypothesize that elongated polyQ-AR in muscle cells sequester different proteins from functional cycles of cells and make the intranuclear fibrillar aggregates. The presence of these aggregates cause cell loss and toxicity in muscle cells, while in spinal cord, due to slow progression of the disease, the cell loss will be observed in a later stage of the disease.

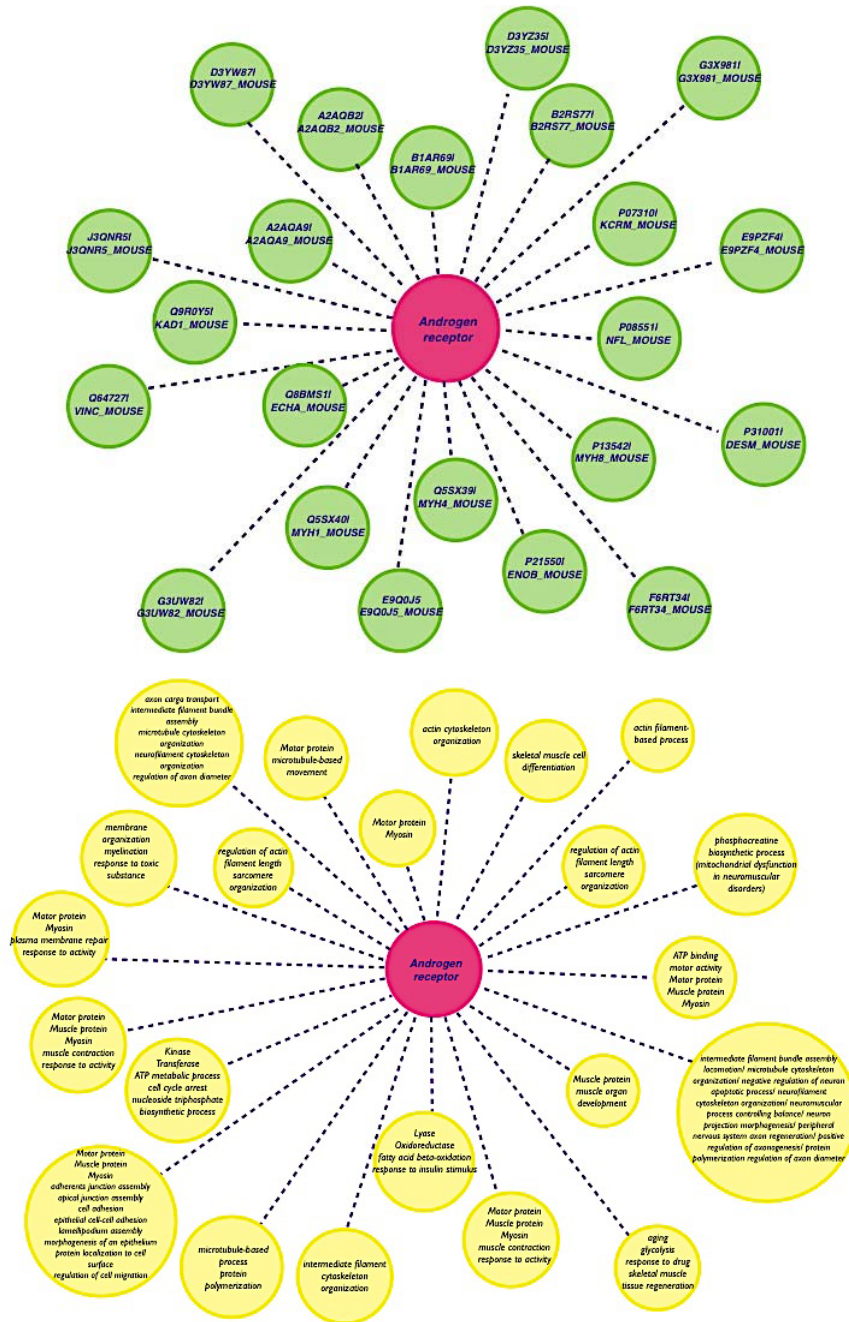


Figure 6-39: An overview on experimental data on the AR interacting protein (upper) and their biological functions (lower panel) involved in muscle degeneration procedure of SBMA male transgenic animals. Proteins with high or low abundance were selected from three different age groups of 97Q young, old and very old male mice. These proteins are mainly involved in muscle function. The details of this figure are shown as tables in the Appendix.

6.3. Summary

Transgenic animals of SBMA with an elongated polyQ (97Q) have been already characterized (92). AR 97Q males exhibit hormone and age dependent weakness beginning at 10 weeks of age (314).

To summarize our main findings, it is to mention that we characterized the total tissue homogenates of two main tissues (muscle and spinal cord) involved in SBMA. We demonstrated that SDS-resistant aggregated species in stacking gels, positive for H280 antibody, are only present in 97Q animals while no 24Q mice revealed presence of any of these species. Comparing the spinal cord and muscle tissue homogenates of 97Q animals showed that less SDS resistant aggregates present in spinal cord compared to muscles using specific AR N-terminal antibody. Beside this validation that we have less aggregation in spinal cord compared to muscle, this data proves that the aggregated species in muscles comprise mainly AR-NTD while spinal cord tissue homogenates were positively stained by C-terminal antibody (presence of LBD).

We did a time-lapse tissue homogenate characterization of 97Q animals with different ages and our results showed that a combination of WB analysis and various high-resolution characterization demonstrates a correlation in amounts of aggregates with aging.

For the first time by using AFM and TEM, two distinct Seprion captured- species of fibrillar aggregates in skeletal muscle and pre- fibrillar aggregated species in spinal cord tissues were visualized. It also showed that while 97Q males aging, there is a direct correlation between age and amount of aggregating species. The animals show an increase in the number of fibrillar species in muscle also in the length of fibrils in muscles tissue, as well the amount of aggregating protein in spinal cord increases. Not only the structures, which differ, but also the aggregates from different origins showed different cytotoxicity in cell toxicity assays.

Additionally, among different tissue homogenates of transgenic and non-transgenic animals, only the muscle tissue extracted caused cell toxicity. Finally, thank to our proteomics studies, we identified some critical proteins in tissue homogenates of SBMA 97Q mice that showed significant changes in their abundance as the disease progressed. We assume that these significant changes in the level of these proteins (either decrease or increase) are a direct indication of involvement of these proteins in progression of the disease and pathology. Further investigations on role of each protein and their structural propensities are inevitable.



PART V.

7. Discussion

Discussion

7. Discussion

7.1. Structural characterization of N-terminal domain of AR by Nuclear Magnetic Resonance (NMR)

The polyQ tract in AR is helical

Characterizing the structural and dynamical properties of polyQ tracts is crucial for understanding why their elongation causes protein aggregation and the molecular basis of polyQ diseases. However, they are generally considered very challenging targets for conventional structural biology approaches such as X-ray crystallography and NMR in solution. This is due, in addition to their low solubility, to their high propensity to be intrinsically disordered (ID), which in general precludes crystallization, and to the highly repetitive nature of their sequences, which can render the assignment of their NMR resonances very challenging due to spectral overlap. A number of pioneering studies attempted to determine the structure of polyQ tracts but, to date, it has not been possible to report on the structure of a one such tract in its native context and without fusing it to a solubilizing moiety such as the Glutathione-S-transferase (315).

To accomplish this and, most importantly, to determine how the structural properties of the protein are modified when the tract is elongated, we used relatively sophisticated sample preparation methods and a battery of newly developed solution NMR techniques such as direct ^{13}C detection experiments and NUS that are well suited to the characterization of ID proteins. The results that we obtained, allowed, for the first time, to fully assign the backbone resonances of the protein, including the polyQ tract, and to derive from the assignments the secondary structure propensity of every residue belonging to it. Contrary to what is the case for the vast majority of such tracts studied to date, our results indicate that the polyQ tract of the NTD of AR has a strong propensity to adopt an α -helical secondary structure (315–317).

When the tract is very short, as is the case for 4Q, the helicity is modest ($\text{SSP} \approx 0.3$) but when the length of the tract reaches values in the physiological range, as is the case for 25Q, the helicity is essentially complete for the first residues of the tract ($\text{SSP} \approx 1$) and then decays along the sequence until it becomes comparable to that of the rest of the residues of the protein ($\text{SSP} \approx 0.2$) (Fig. 4-6). The wide range of helical propensities observed in the tract is likely responsible for the relatively wide dispersion of ^1H , ^{15}N and, especially, ^{13}C chemical shifts that is apparent in the 2D spectra shown in Figure 4-5 and that was instrumental for obtaining the essentially complete set of assignments that is the basis of this work.

Detailed inspection of both the spectra shown in Figure 4-5B and the SSP and ^{15}N R_2 values shown in Figure 4-6B reveals that the induction of helicity due to elongation is not limited to the tract but affects the residues flanking it at its N-

terminus. The four Leu residues of the motif ⁵⁵LLLL⁵⁸ that precede the tract undergo very substantial ¹H, ¹⁵N and ¹³C chemical shift changes upon elongation and increase their helical propensity as much as Gln residues 59 to 62; by contrast the chemical shifts and, therefore, the structural properties of the rest of residues of the protein remain unaltered. This result indicates a certain degree of cooperativity in the conformational transition caused by elongation and strongly suggests that the ⁵⁵LLLL⁵⁸ motif nucleates the helical conformation of the polyQ tract. This hypothesis is supported by a substantial body of literature describing polyQ sequences as random coils (318, 319) and by the high helical propensity of Leu residues (320).

The regions flanking polyQ tracts have been the focus of much attention because their presence alters the kinetics of aggregation by facilitating nucleation or, as has been suggested, by modifying the structural properties of the tract (321–323). Here we show a rather extreme case of the latter as the presence of a LLLL motif N-terminal to a polyQ tract appears to nucleate a helical conformation in the tract to the extent that a substantial part of it forms a fully formed helix. It is interesting to speculate about the evolutionary advantage provided by the LLLL motif: although it is possible that it plays a functional role, it can also act as a mechanism to prevent or at least slow down the rate of aggregation by folding the tract into a helical conformation, which has a intrinsically low propensity to aggregate (324).

Oligomerization is triggered by a motif distant from the polyQ tract

Studies of the mechanism of aggregation of proteins with polyQ tracts such as AR mainly relied on comparing the aggregation rates of synthetic peptides with tracts of different lengths and after introducing various changes in the sequence of the flanking regions (325, 326) These studies have showed a strong correlation between the length of the polyQ and the rate of aggregation of the corresponding peptide (327) and, importantly, have suggested that the flanking regions can play a key role in the early stages of the mechanism of aggregation by becoming involved in intermolecular interactions that stabilize on-pathway oligomers (328). Our results provide new insights into how 25Q self assembles by directly reporting on the specific residues that undergo changes in structure and dynamics as oligomerization progresses which is, as far as we are aware, unprecedented in the study of protein oligomerization.

The results that we obtained with the urea incubation followed by dialysis (UID) sample indicate that under conditions where oligomerization is very slow, 25Q is in equilibrium with an oligomeric form of the protein stabilized by inter-molecular interactions involving residues in the motif ²²GAFQNLFQ²⁹ (Fig. 4-11). This motif is in a region of sequence of low helical propensity but high AABUF (Fig. 4-1), suggesting that the interactions that stabilize the oligomers that form earliest are hydrophobic in nature. Given that the equilibrium between monomeric 25Q and these oligomers is relatively fast in the NMR timescale, it is unlikely that 25Q

undergoes substantial structural changes of the type that are observed in the SDP sample discussed below, indicating that they are disordered.

We characterized the residues that experience chemical shift perturbations during the oligomerization of 25Q in a SDP sample (Fig. 4-9). The region of sequence where most of such residues are found is the motif $^{30}\text{SVREVIQ}^{36}$, of predicted high helicity, that is part of the region $^{22}\text{GAFQNLQSVREVIQ}^{36}$ and is directly adjacent to the motif $^{22}\text{GAFQNLQ}^{29}$. That the species with different chemical shifts in the $^{30}\text{SVREVIQ}^{36}$ motif are in slow exchange with monomeric 25Q indicates that they correspond to well-defined oligomeric species of a size that is compatible with their characterization by solution NMR. We conclude that the second step in the oligomerization of 25Q is the formation of oligomers stabilized by inter-molecular interactions involving mainly residues in the motif $^{30}\text{SVREVIQ}^{36}$ (Fig. 4.9B).

The chemical shift perturbations observed in the $^{30}\text{SVREVIQ}^{36}$ occur concomitantly with decreases in signal intensity in the resonances of monomeric 25Q (Fig. 4.9B). It is important to clarify that this observation is specific to a subset of the resonances, as a substantial number of residues of 25Q (residues 1-10 and 90-120 in Fig. 4-9B) do not experience a decrease in signal intensity. This observation confirms that the phenomenon reported in Fig. 4-9B is not due to the aggregation of 25Q but, rather, to the changes occurring in the structure and the dynamics of the protein as it oligomerizes (Fig. 4-10B).

By using the decreases in signal intensity as reporters of structure formation in the oligomers, we can determine the order in which the various regions of sequence of 25Q become partially folded. Based on this principle, we can conclude that the polyQ tract and the regions of sequence immediately flanking undergo conformational changes only after the region $^{22}\text{GAFQNLQSVREVIQ}^{36}$ becomes partially structured. This is clearly illustrated by the intensities reported at $t = 2\text{d}$ in Figure 3B, which are substantially decreased in the region $^{22}\text{GAFQNLQSVREVIQ}^{36}$ but remain essentially unchanged in the polyQ tract. We thus conclude that the third step in the oligomerization of 25Q is a conformational change in the polyQ region and that it occurs in an on-pathway oligomer stabilized by intermolecular interactions involving the region $^{22}\text{GAFQNLQSVREVIQ}^{36}$ and, especially, the motif $^{30}\text{SVREVIQ}^{36}$ (Fig. 4.9B).

Implications for the early stages of AR aggregation in SBMA

The oligomerization mechanism described above for 25Q is similar to the one put forward for exon 1 of the Htt on the basis of the effect of mutations in the flanking regions on the rate of monomer depletion (323). For this case it was suggested that the interactions between flanking regions that stabilize the oligomers are equivalent to those occurring in coiled coils, which are frequent in transcription factors. This was based on predictions by bioinformatics tools as well as on the observation that the main flanking region in Htt has helical character according to both CD and NMR spectroscopies (45, 329, 330).

An analysis of the sequence of 25Q in the single region that experiences the earliest and largest decreases in signal intensity during oligomerization as well as the largest chemical shift perturbations indicates that it has some of the features expected in a sequence that forms coiled coil such as the presence of hydrophobic residues in positions *a* and *d* of two consecutive imperfect heptad repeats starting at positions F24 and V31. This, together with the observation that the two motifs that most decrease in intensity as oligomerization progresses have some predicted helical propensity strongly suggest that, as proposed for Htt, the on-pathway oligomers that form during the aggregation of 25Q are stabilized by coiled coil interactions established by partially helical motifs.

The results that we present in this work have implications for our understanding of the mechanism of oligomerization *in vivo* and for the identification of therapeutic strategies for SBMA. As already mentioned, the motif that establishes the earliest inter-molecular interaction in 25Q, ²²GAFQNL²⁸, binds to AF-2 in the LBD after activation of AR by androgens. Our identification of this motif as the site of initiation of aggregation can provide a simple explanation of why cleavage of AR by Caspase-3 facilitates oligomerization and the onset of SBMA. Proteolysis by Caspase-3 produces two AR fragments, one containing the ²²GAFQNL²⁸ motif and another one containing the LBD. Although cleavage does not physically prevent the motif and the LBD from interacting, it is certain to substantially decrease the affinity between these two moieties by increasing the entropic cost of binding. This decrease in affinity will inevitably lead to an increase in the concentration of free N-terminal fragment and, given the properties of the motif, to an acceleration of its oligomerization and the onset of SBMA. Finally, we note that our work clearly identifies the region ²²GAFQNL²⁸QSVREVIQ³⁶ and the ²²GAFQNL²⁸ motif in particular as therapeutic targets for the development of treatments for SBMA (Fig. 4-10B).

7.2. Mapping of Hsps binding to AR

There has been a great interest in understanding how the molecular chaperones DnaK (Hsp70) and DnaJ (Hsp40) interact with substrates to assist their folding. Although many investigations have been performed to explore the role of Hsps in pathology of SBMA (39, 160, 163, 176, 180), there is yet not even one high-resolution information on the direct interaction between Hsps and Wt AR. Many groups have been studying the involvement of Hsp machinery in different neurodegenerative diseases' models in cells and Transgenic animals (39, 179, 238, 331). However, none was able to validate the properties, kinetics or motifs involved in amyloid proteins-Hsp interaction, except for the recent data on tau protein interaction with Hsp90 (265).

In this work, thanks to our parallel solution state NMR approach, which helped us to fully characterize the AR N-terminal domain (Chapter 4), we indicated for the first time that both Hsp40 and Hsp72 bind to AR.

Due to our primary interest on the role of Hsp40 and Hsp72 with regard to SBMA pathology, instead of a full-length AR N-terminal domain, we studied the biologically relevant N-terminally truncated polyQ AR with two different sizes of the polyQ (Fig. 4-1A and 4-1C). Interestingly enough, the mapped regions on the short construct of N-terminal domain of AR made it possible for us to have a clear view of how Hsps play their protective role on AR.

Our studies showed that both Hsp72 and Hsp40 bind to three motifs on AR, however with different binding affinity. Our results demonstrated that Hsp72 mainly binds to specific residues of ²⁴FQNLFQSVREVIQ³⁶ with a main core of interaction on ²⁴FQNLF²⁸ region in both 4Q and 25Q, while Hsp40 binds with less specificity to two additional regions with also high hydrophobicity, a typical trigger for chaperon interaction (332) (Fig. 4-1B). These motifs are ⁵⁵LLLL⁵⁸ and ¹⁰⁶GPTGYLVL¹¹⁴. A higher affinity for ⁵⁵LLLL⁵⁸ region has been shown in the 4Q construct compared to 25Q.

It is well accepted in the Hsp field that most of the Hsps bind to the least structured parts of a protein and manage the folding of the proteins. Recent data showed that most of these Hsp's binding pockets are composed of hydrophobic amino acids (265) and are therefore prone to form intermolecular interactions. In case of the AR, the very same properties were also observed for binding epitopes of both Hsp40 and Hsp72. The three main binding motifs on AR are exactly those regions of sequence with highest AABUF (Fig. 4-1).

The ²⁴FQNLF²⁸ region and its proximate flanking amino acids are (1) the initiation of the oligomerization in this AR construct (Fig. 4-10B), (2) this region of the protein in Wt AR has well identified interaction with C-terminal domain of the protein and is responsible for the well-characterized N/C intramolecular interaction (108). Therefore, from an evolutionary point of view, it is necessary for the cells to protect this region of protein, since the high intramolecular interaction propensities of the motif might result either to aggregation or to inappropriate protein-protein interactions.

To summarize, while AR is in cytoplasm and has no transcriptional activity, Hsp72 protects ²⁴FQNLF²⁸ region from aggregation and non-functional protein-protein interactions. In presence of DHT and its binding to ligand binding domain (LBD), the higher affinity of C-terminal region to ²⁴FQNLF²⁸ motif, makes the Hsps leave and prepare the protein for translocation to the nucleus. Hereby, we propose that LBD in full-length AR functions as internal chaperone to compensate the absence of the Hsps, while protein is translocating to nucleus. Our findings highlight the therapeutic potential of allosteric regulators of Hsp70 and Hsp40 and provide new insights into the role of the chaperone machinery in protein quality control.

7.3. Characterization of AR aggregates in SBMA transgenic mouse model

To date, studies of the patients with SBMA and other polyQ diseases or respective mouse models were not able to provide any structural information, which explains the toxicity of polyQ aggregates and their relation to the mechanism and pathology of the disease. Previous studies illustrated that the intranuclear aggregates of AR are the main entities in the polyQ disease mechanism (15, 220), while no thorough explanations of how this protein can have toxic properties and how polyQ aggregates can attenuate the cell cascades activities, have been provided. Therefore, it has not been feasible to find a correlation between the presence of specific species and their toxicity as the disease progresses.

Previous work from Prof. Sobue's laboratory has detected AR aggregates in the stacking gels of WB analysis (33, 94). Different groups have been doing biochemical characterizations of intermediate and late stage aggregated species in the disease models (87, 92); however, no studies characterized the morphology of different toxic species in an age-dependent manner in SBMA transgenic mice. Structural studies are limited only to AR aggregates prepared *in vitro* (333). There is still no high-resolution structural information about the SDS-resistant aggregated species of AR.

In this work, we presented the first evidence that structurally speaking, there are two thoroughly distinct species of AR toxic protein that present in spinal cord and skeletal muscle tissue (Figs. 6-13 and 6-15). Our high-resolution structural information from AFM in a time resolved fashion and various microscopy methods, illustrated that there is a structure-activity relationship between the presence of unique fibrillar species and their toxicity as well as the disease phenotype. Our studies link the structure to the onset and progression of the disease. The aggregates captured from muscle tissues are fibrillar and have a well-ordered structure and length, which is interestingly in the size order of the *in vitro* produced fibrillar species of N-terminally truncated AR (Fig. 6-14), while the captured material from spinal cord comprises mainly the pre-fibrillar globular species. The aggregates in muscles increase in their length and number as the disease progress, whereas the non-fibrillar spinal cord aggregates don't change significantly by aging. These data suggest that muscle degeneration is prior to the motor neuron degenerations in spinal cord and if any propagation of aggregated species could potentially happen, then it will be from muscle to spinal cord and not vice versa.

The identification of the main toxic protein and its involvement in the mechanism of the disease can hold great promises in the future view of the SBMA field. Due to the presence of more than one toxic species in SBMA transgenic animals, an anti-aggregatory approach needs to be directed more to specific structural variants and as well to specific tissue.

Current knowledge obtained from our experiments suggests that involvement of skeletal muscles as the primary event in the phenotype of the positive 97Q mice could open new approaches and views toward a therapy for SBMA. With new available data in the field, further investigations could explain the whole mechanism of the pathology in skeletal muscle tissue and consequently in the central nervous system. This might also be the time to bring up a new definition for SBMA, changing from motor neuron disease to a muscle degenerative disease.



8. Concluding remarks

8.1. What are the next steps toward a treatment for SBMA?

8. Concluding remarks

The experimental work reported in this thesis provides a structural and biophysical perspective that contributes to expanding the current knowledge about the mechanism of the aggregation formation of AR and its toxicity in the Spinal Bulbar Muscular Atrophy (SBMA).

In three projects that I was involved in during my PhD, I tried to bring the *in vitro* data to the *in vivo* data and provide a complete view on both structural and molecular mechanism of SBMA pathology.

1- Our *in vitro* work with solution state NMR as the main tool, made it possible, for the first time, to assign a polyQ protein. Assignment of all the residues of the N-terminal truncated AR construct, we obtained information on the structure and the dynamics of this protein. The results that we present in this work have implications for our understanding of the mechanism of oligomerization *in vivo* and for the identification of therapeutic strategies for SBMA. Our work clearly identifies the region ²²GAFQNLFQSVREVIQ³⁶ and the ²²GAFQNLF²⁸ motif in particular as therapeutic targets for the development of treatments for SBMA.

2- In a parallel work on importance of the Hsps interacting with polyQ proteins, for the first time, we identified the binding epitopes of Hsp72 and Hsp40 on the AR N-terminal domain. These findings provided us with high-resolution binding site information, which was validated as well with cell experiments. These findings may provide the Hsps and the polyQ diseases' fields with the opportunity to use biophysical methods as screening tools in study of the small molecule therapeutics and drug development based on Hsps modulators.

3- Our *in vivo* work with transgenic mouse model of SBMA revealed a possible explanation for the mechanism of the disease and the involvement of skeletal muscle cells preferred over motor neurons, in the progression of the disease. Our data demonstrate the toxicity of AR protein in the context of *in vivo* system for the first time.

The general conclusions arising from these findings can be summarized as following:

1. The N-terminal domain of AR is predicated as an intrinsically disordered protein (IDP), but our studies showed that this domain comprises small motifs with secondary structure propensities. We identified two critical regions in AR that are not only important in SBMA pathology, but also in the intrinsic function of AR and its role in the mechanism of prostate cancer. Based on the structural characterization prediction data, ²⁴FQNLF²⁸ does not have a well-defined secondary structure; nevertheless, this does not mean that it does not get any secondary structure upon binding. In our study, this region was indicated as the initiation site for oligomerization of the AR. Therefore, we propose that, this

region is also critical for the intrinsic function of AR dimer formation. Moreover, the region of interest ⁵⁵LLLLQQQQ⁶², which is just flanking to the elongated polyQ, adopts an α -helical structure, which initiates from ⁵⁵LLLL⁵⁸ part in flanking regions of polyQ.

2. Both Hsp72 and Hsp40 bind to AR in a sequence specific fashion. They both share the ²⁴FQNL²⁸ as the binding motif, while Hsp40 has two more binding regions. Hsp40 binds additionally to ⁵⁵LLLLQQQQ⁶² and ¹⁰⁶GPTGYLVLD¹¹⁴ regions in both 25Q and 4Q. These distinct sequences comprise specific structural properties due to the presence of highly hydrophobic residues, which triggers the chaperones interaction. It is of high emphasis for cells to protect these three less flexible hydrophobic regions by binding to Hsps, since otherwise; the intrinsic function of AR would be diminished, as the protein could start to oligomerize and subsequently aggregate.

3. Structural differences in *in vivo* captured aggregates from 97Q animal's tissue homogenates suggest the mechanism of the disease and involvement of muscle tissue as the main pathological tissue during the disease's progression. We hypothesize that the fibrillar species in skeletal muscles primarily sequester many proteins into aggregates and attenuate the cellular functions. Nevertheless, in later stages of the disease, the intranuclear aggregates will also form in the spinal cord.

8.1. What are the next steps toward a treatment for SBMA?

Kennedy's disease (spinal and bulbar muscular atrophy) is a neuromuscular disorder characterized by slowly progressive muscle weakness and atrophy, with degeneration of primary motor neurons in the spinal cord and brainstem of affected individuals (17). Traditionally, SBMA has been viewed as a cell-autonomous, primary motor neuron disease (98). The neuronal death in SBMA and other polyQ diseases is at least in part due to a toxic gain-of function in the mutant protein. In addition to the toxic effects of the AR, loss of normal receptor function also contributes to the disease phenotype, with some patients having symptoms such as breast enlargement and reduced fertility (15).

And therefore, an approach aimed at reducing mutant protein levels without exacerbating the effects of androgen deficiency, holds great promise as a potential treatment for SBMA (94, 334).

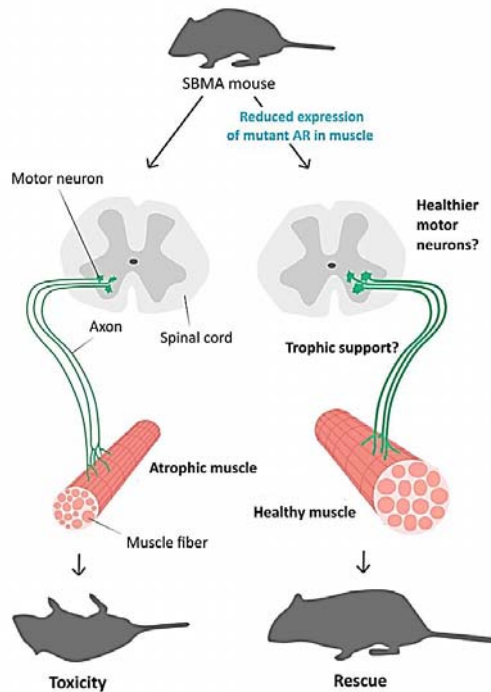


Figure 8-1. Reducing Mutant AR Expression in Muscle Has Beneficial Effects in SBMA Mouse Models (Adapted from (90))

A number of different treatments have been shown to protect against expanded polyQ AR toxicity in SBMA animal models (see the introduction); nevertheless, they have so far failed to result in an effective treatment in human clinical trials (270, 335, 336). This failure can be partly due to complications in preclinical research, such as a lack of blinding and starting treatment before the onset of manifestations of the disease, which is difficult to do in clinical trials.

The development of therapeutics would also benefit from an improved understanding of the mechanism of muscle and motor neuron degeneration in SBMA. Non-neuronal cells such as glial cells and muscle probably play a critical role in the pathogenesis of motor neuron diseases (90). Appreciation of this role not only gives a better insight into the disease mechanism, but also opens new treatment strategies, which are particularly desirable for these disorders. For example, cells outside the CNS may play a primary role in spinal muscular atrophy (SMA). Peripheral targeting of antisense oligonucleotides (ASOs) to restore the expression of survival motor neuron protein, the deficiency of which causes SMA, robustly ameliorates the disease manifestations in SMA mice (337). Moreover, muscle-specific conditional rescue in SMA mice leads to a significant improvement in weight, survival, and motor behavior (338). Reports by Cortes et al., 2014 in *Neuron* (88) and Lieberman et al., 2014 in *Cell* (89) challenged the traditional view of SBMA as a primary motor neuron disease. These studies establish the muscle as

the main site of mutant AR toxicity and suggested that targeting mutant protein expression in this tissue is the approach for treating the SBMA (Fig. 8-1).

Several lines of evidence from previous studies support a primary contribution of skeletal muscle in the disease pathogenesis: (1) muscle biopsies of SBMA patients show features of both denervation and myofiber degeneration (10) ; (2) knock-in mice expressing polyQ- expanded AR develop early signs of myopathy with little or no motor neuron loss (93); (3) muscle-specific overexpression of Wt and non-expanded AR in mice is sufficient to produce a SBMA-like neuromuscular disease (98) ; and (4) genetic overexpression of muscle specific IGF-1 or peripheral IGF-1 administration has been shown to reduce SBMA phenotype in the transgenic mice models (87, 339). Based on the recent studies in this field, selective suppression of the AR transgene in skeletal muscle is the new approach, which the whole field is moving towards. It has been shown that, muscle-specific removal of AR, increased survival, suppressed weight loss and weakness and increased the diameter of motor axons (89). Consequently, this study demonstrated a primary role of skeletal muscle in SBMA pathogenesis in these mice and thus justifies the AR gene silencing in muscle tissues as a potential disease-modifying strategy for patients (90).

By targeted reduction of mutant AR using ASO technology, Lieberman et al., 2014, took this approach one step closer to the clinic. In various mouse models of SBMA, ASOs were delivered subcutaneously, before or at the time of disease onset and they have turned out to significantly suppress AR expression in skeletal muscle, but not spinal cord. The ASO treatment rescued weight loss, muscle weakness, abnormal gene expression and lethality in the mice, without altering testosterone levels.

In particular, AR suppression by ASOs was maintained up to 10 weeks after terminating of the treatment. Decreasing the AR expression level in the spinal cord was also achieved through intraventricular delivery of ASO, but surprisingly did not modify disease manifestation, indicating that the main problem in these mice is not in the CNS and that the best way to improve it is with peripheral rather than central treatment strategies.

There are practical advantages in targeting the periphery for treatment over the delivery to the CNS. This approach would limit the known adverse effects of androgen reduction therapy related to AR reduction in the CNS, such as decreased libido. Moreover, skeletal muscle is easily accessible; it can be biopsied for pharmacodynamics measures of drug effects; unlike the spinal cord.

Based on our structural and biophysical data and the recent publications, it is tempting to think of SBMA as a muscle disease rather than a motor neuron disease and therefore we have to try to mainly address this tissue in our therapeutic strategies.

PART VI.

9. Appendix

- 9.1. AR sequence and residue numbering used in this thesis, according to Uniprot
- 9.2. HSQC assignment of AR-4Q
- 9.3. CON assignment of AR-4Q
- 9.4. HSQC assignment of AR-25Q
- 9.5. pDEST-HisMBP vector
- 9.6. Ni⁺² affinity chromatography
- 9.7. Size exclusion chromatography
- 9.8. Reverse Ni⁺² chromatography
- 9.9. Mass spectrometry analysis of 4Q and 25Q

9. Appendix

9.1. AR sequence and residue numbering used in this thesis, according to Uniprot

The AR sequence deposited in the Uniprot database (entry P10275, isoform 1), corresponding to 919 residues for the full-length protein, and with 21 glutamine residues in the polymeric glutamine stretch (between residues 58 and 78) and 24 glycine residues in the polymorphic glycine stretch (between residues 449 and 472). The residue numbering corresponding to this AR sequence is used throughout the thesis. But as the construct of A-25Q and AR-4Q both after TEV cleavage site have a G before the first M, the whole numbering of residue has changed from i to $i+1$.

10 20 30 40 50 60
 MEVQLGLGRV YPRPPSKTYR GAFQNLFSV REVIQNPGR HPEAASAAPP GASLLLLQQQ

70 80 90 100 110 120
 QQQQQQQQQ QQQQQQQQET SPRQQQQQQG EDGSPQAHRR GPTGYLVLDE EQQPSQPQSA

130 140 150 160 170 180
 LECHPERGCV PEPGAAVAAS KGLPQQLPAP PDEDDSAAPS TLLSLLGPTFF GLSSCSADLK

190 200 210 220 230 240
 DILSEASTMQ LLQQQQQFAV SEGSSSGRAR EASGAPTSSK DNYLGGTSTI SDNAKELCKA

250 260 270 280 290 300
 VSVSMGLGVĒ ALEHLSPEQ LRGDCMYAPL LGVPPAVRPT PCAPLAECKG SLLDDSAAGS

310 320 330 340 350 360
 TEDTAFYSPĒ KGGYTKGLEG ESLGCSGSAĒ AGSSGTLELP STLSLYKSGA LDEAAAYQSR

370 380 390 400 410 420
 DYYNFPLALĒ GPPPPPPPH PHARIKENP LDYGSAAAAĒ AAQCRYGLĒ SLHGAGAAGP

430 440 450 460 470 480
 GSGSPSAAAS SSWHTLFTAĒ EGQLYGPCGG GGGGGGGGGG GGGGGGGGGG GGEAGAVAPY

490 500 510 520 530 540
 GYTRPPQGLĒ GQESDFTAPD VWYPGGMVSR VPYPSPTCVK SEMGPWMSY SGPYGDMRLE

550 560 570 580 590 600
 TARDHVLPID YYFPPQKTCL ICGDEASGCH YGALTCGSCK VFKRAAEGK QKYLCA SRND

610 620 630 640 650 660
 CTIDKFRKRN CPSCRLRKCY EAGMTLGARK LKKLGNLKLQ EEGEASSTTS PTEETTQKLT

670 680 690 700 710 720
 VSHIEGYECQ PIFLNVLEAI EPGVVCAGHD NNQPD SFAAL LSSLNELGER QLVHVVKWAK

730 740 750 760 770 780
 ALPGFRNLHV DDQMAVIQYS WMGLMVFAMG WRSFTNVNSR MLYFAPDLVF NEYRMHKSRM

790 800 810 820 830 840
 YSQCVRMRHL SQEFGWLQIT PQEFLCMKAL LLFSIIPVDG LKNQKFFDEL RMNYIKELDR

850 860 870 880 890 900
 IIACKRKNPT SCSRRFYQLT KLLDSVQPIĒ RELHQFTFDL LIKSHMVSVD FPEMMAFIIS

910
 VQVPKILSGK VKPIYFHTQ

9.2. HSQC assignment of AR-4Q

The data were recorded at 425 μ M 4Q, in 20mM sodium phosphate buffer and 1mM TCEP, at pH 7.4 and 278 K.

Assignment	w1	w2
E3N-E3H	123.227	8.877
V4N-V4H	122.468	8.385
Q5N-Q5H	124.838	8.603
L6N-L6H	124.356	8.53
G7N-G7H	109.591	8.559
L8N-L8H	121.63	8.284
G9N-G9H	109.702	8.602
R10N-R10H	120.772	8.148
V11N-V11H	122.184	8.265
Y12N-Y12H	126.353	8.598
S17N-S17H	116.709	8.599
K18N-K18H	123.648	8.656
T19N-T19H	115.13	8.244
Y20N-Y20H	123.769	8.394
R21N-R21H	124.649	8.478
G22N-G22H	109.266	7.775
A23N-A23H	123.411	8.283
F24N-F24H	119.402	8.373
Q25N-Q25H	121.791	8.293
N26N-N26H	119.718	8.473
L27N-L27H	122.501	8.262
F28N-F28H	120.174	8.285
Q29N-Q29H	121.793	8.242
S30N-S30H	117.696	8.442

9. Appendix

V31N-V31H	122.156	8.321
R32N-R32H	125.038	8.474
E33N-E33H	123.027	8.505
V34N-V34H	123.151	8.417
I35N-I35H	126.385	8.422
Q36N-Q36H	125.568	8.661
N37N-N37H	121.461	8.707
G39N-G39H	109.282	8.381
R41N-R41H	121.648	8.587
H42N-H42H	122.535	8.516
E44N-E44H	121.581	8.936
A45N-A45H	125.565	8.456
A46N-A46H	123.729	8.462
S47N-S47H	115.27	8.36
A48N-A48H	126.234	8.385
A49N-A49H	125.232	8.365
G52N-G52H	109.604	8.646
A53N-A53H	123.761	8.235
S54N-S54H	114.838	8.424
L55N-L55H	124.433	8.298
L56N-L56H	121.274	8.068
L57N-L57H	121.565	8.006
L58N-L58H	121.62	8.087
Q59N-Q59H	120.115	8.315
Q60N-Q60H	120.769	8.395
Q61N-Q61H	121.062	8.462
Q62N-Q62H	121.123	8.453
E63N-E63H	122.226	8.563

T64N-T64H	115.176	8.348
S65N-S65H	119.8	8.434
R67N-R67H	121.45	8.551
Q68N-Q68H	121.778	8.547
Q69N-Q69H	122.019	8.609
Q70N-Q70H	121.864	8.591
Q71N-Q71H	122.072	8.595
Q72N-Q72H	122.258	8.656
Q73N-Q73H	122.271	8.675
G74N-G74H	110.619	8.623
E75N-E75H	120.847	8.519
D76N-D76H	121.32	8.585
G77N-G77H	109.932	8.496
S78N-S78H	117.401	8.287
Q80N-Q80H	120.117	8.539
A81N-A81H	124.786	8.342
H82N-H82H	118.219	8.297
R83N-R83H	122.468	8.377
R84N-R84H	122.678	8.618
G85N-G85H	110.32	8.454
T87N-T87H	112.083	8.336
G88N-G88H	110.619	8.339
Y89N-Y89H	120.192	8.104
L90N-L90H	124.252	8.346
V91N-V91H	122.995	8.301
L92N-L92H	126.974	8.501
D93N-D93H	121.76	8.454
E94N-E94H	121.249	8.541

9. Appendix

E95N-E95H	121.733	8.549
Q96N-Q96H	121.418	8.473
Q97N-Q97H	122.876	8.596
S99N-S99H	117	8.628
Q100N-Q100H	122.984	8.542
Q102N-Q102H	121.334	8.754
S103N-S103H	117.536	8.521
A104N-A104H	126.322	8.54
L105N-L105H	120.778	8.22
E106N-E106H	121.411	8.346
C107N-C107H	120.601	8.37
H108N-H108H	122.964	8.665
E110N-E110H	121.522	9.11
R111N-R111H	122.364	8.468
G112N-G112H	109.948	8.548
C113N-C113H	119.058	8.354
V114N-V114H	124.722	8.451
E116N-E116H	123.053	8.639
G118N-G118H	109.681	8.688
A119N-A119H	123.782	8.123
A120N-A120H	123.603	8.439
V121N-V121H	120.214	8.216
A122N-A122H	128.15	8.492
A123N-A123H	123.635	8.435
S124N-S124H	115.362	8.378
K125N-K125H	123.419	8.449
G126N-G126H	109.677	8.442
L127N-L127H	123.047	8.26

Q129N-Q129H	120.945	8.62
Q130N-Q130H	122.553	8.584
L131N-L131H	126.185	8.579

9.3. CON assignment of AR-4Q

The data were recorded at 425 μ M 4Q, in 20mM sodium phosphate buffer and 1mM TCEP, at pH 7.4 and 278 K.

Assignment	w1	w2
N26ND2-N26CG	113.121	176.963
N37ND2-N37CG	113.721	177.035
M2N-G1C	120.009	170.778
E3N-M2C	123.227	176.236
V4N-E3C	122.468	176.616
Q5N-V4C	124.838	176.224
L6N-Q5C	124.356	175.961
G7N-L6C	109.591	178.064
L8N-G7C	121.63	174.388
G9N-L8C	109.702	178.18
R10N-G9C	120.772	173.875
V11N-R10C	122.184	175.715
Y12N-V11C	126.353	175.579
P13N-Y12C	136.644	173.751
P15N-R14C	138.792	174.017
P16N-P15C	135.599	174.671
S17N-P16C	116.709	176.977
K18N-S17C	123.648	174.847
T19N-K18C	115.13	176.637

9. Appendix

Y20N-T19C	123.769	174.076
R21N-Y20C	124.649	175.624
G22N-R21C	109.266	176.259
A23N-G22C	123.411	173.555
F24N-A23C	119.402	177.812
Q25N-F24C	121.791	175.776
N26N-Q25C	119.718	175.213
L27N-N26C	122.501	175.218
F28N-L27C	120.174	177.218
Q29N-F28C	121.793	175.655
S30N-Q29C	117.696	175.543
V31N-S30C	122.156	174.579
R32N-V31C	125.038	176.138
E33N-R32C	123.027	176.141
V34N-E33C	123.151	176.202
I35N-V34C	126.385	176.134
Q36N-I35C	125.568	176.166
N37N-Q36C	121.461	175.324
P38N-N37C	136.731	173.234
G39N-P38C	109.282	177.088
P40N-G39C	134.299	171.619
R41N-P40C	121.648	177.024
H42N-R41C	122.535	176.021
P43N-H42C	137.591	173.733
E44N-P43C	121.581	177.074
A45N-E44C	125.565	176.472
A46N-A45C	123.729	177.689
S47N-A46C	115.27	177.974

A48N-S47C	126.234	174.157
A49N-A48C	125.232	177.052
P50N-A49C	137.206	175.004
P51N-P50C	135.956	174.895
G52N-P51C	109.604	177.816
A53N-G52C	123.761	174.355
S54N-A53C	114.838	178.467
L55N-S54C	124.433	175.28
L56N-L55C	121.274	178.019
L57N-L56C	121.565	177.933
L58N-L57C	121.62	178.052
Q59N-L58C	120.115	178.173
Q60N-Q59C	120.769	176.774
Q61N-Q60C	121.062	176.684
Q62N-Q61C	121.123	176.356
E63N-Q62C	122.226	176.273
T64N-E63C	115.176	176.809
S65N-T64C	119.8	174.506
P66N-S65C	137.959	172.655
R67N-P66C	121.45	177.2
Q68N-R67C	121.778	176.701
Q69N-Q68C	122.019	176.226
Q70N-Q69C	121.864	176.156
Q71N-Q70C	122.072	176.126
Q72N-Q71C	122.258	176.092
Q73N-Q72C	122.271	175.963
G74N-Q73C	110.619	176.502
E75N-G74C	120.847	174.29

9. Appendix

D76N-E75C	121.32	176.619
G77N-D76C	109.932	176.922
S78N-G77C	117.401	174.234
P79N-S78C	138.004	173.009
Q80N-P79C	120.117	177.325
A81N-Q80C	124.786	176.072
H82N-A81C	118.219	177.653
R83N-H82C	122.468	175.378
R84N-R83C	122.678	175.977
G85N-R84C	110.32	176.602
P86N-G85C	134.165	172.295
T87N-P86C	112.083	177.705
G88N-T87C	110.619	175.128
Y89N-G88C	120.192	173.668
L90N-Y89C	124.252	175.432
V91N-L90C	122.995	176.63
L92N-V91C	126.974	175.945
D93N-L92C	121.76	177.011
E94N-D93C	121.249	176.403
E95N-E94C	121.733	176.705
Q96N-E95C	121.418	176.621
Q97N-Q96C	122.876	175.913
P98N-Q97C	137.233	173.992
S99N-P98C	117	176.99
Q100N-S99C	122.984	174.314
P101N-Q100C	137.509	173.979
Q102N-P101C	121.334	177.045
S103N-Q102C	117.536	176.335

A104N-S103C	126.322	174.509
L105N-A104C	120.778	177.887
E106N-L105C	121.411	177.626
C107N-E106C	120.601	176.106
H108N-C107C	122.964	173.984
P109N-H108C	137.613	173.387
E110N-P109C	121.522	177.196
R111N-E110C	122.364	176.838
G112N-R111C	109.948	176.918
C113N-G112C	119.058	173.875
V114N-C113C	124.722	174.566
P115N-V114C	139.681	174.395
E116N-P115C	123.053	176.545
P117N-E116C	138.16	175.015
G118N-P117C	109.681	177.855
O119N-G118C	123.782	174.108
A120N-O119C	123.603	177.863
V121N-A120C	120.214	178.16
A122N-V121C	128.15	176.394
A123N-A122C	123.635	177.806
S124N-A123C	115.362	178.167
K125N-S124C	123.419	174.838
G126N-K125C	109.677	177.014
L127N-G126C	123.047	173.736
P128N-L127C	136.447	175.496
Q129N-P128C	120.945	176.912
Q130N-Q129C	122.553	175.912
L131N-Q130C	126.185	175.673

P132N-L131C	136.103	175.149
A133N-P132C	126.201	176.218
P134N-A133C	137.486	175.095
P135N-P134C	140.302	174.067

9.4. HSQC assignment of AR-25Q

The data were recorded at 375 μ M 25Q, in 20mM sodium phosphate buffer and 1mM TCEP, at pH 7.4 and 278 K.

Assignment	w1	w2
E3N-E3H	123.232	8.861
V4N-V4H	122.491	8.384
Q5N-Q5H	124.851	8.602
L6N-L6H	124.376	8.529
G7N-G7H	109.605	8.555
L8N-L8H	121.635	8.278
G9N-G9H	109.696	8.594
R10N-R10H	120.782	8.141
V11N-V11H	122.202	8.264
Y12N-Y12H	126.373	8.598
S17N-S17H	116.759	8.607
K18N-K18H	123.699	8.66
T19N-T19H	115.182	8.243
Y20N-Y20H	123.824	8.389
R21N-R21H	124.687	8.471
G22N-G22H	109.287	7.742
A23N-A23H	123.411	8.278
F24N-F24H	119.411	8.369

Q25N-Q25H	121.806	8.276
N26N-N26H	119.732	8.472
L27N-L27H	122.533	8.26
F28N-F28H	120.223	8.278
Q29N-Q29H	121.82	8.225
S30N-S30H	117.723	8.433
V31N-V31H	122.199	8.318
R32N-R32H	125.087	8.47
E33N-E33H	123.05	8.503
V34N-V34H	123.151	8.419
I35N-I35H	126.447	8.425
Q36N-Q36H	125.616	8.661
N37N-N37H	121.506	8.71
G39N-G39H	109.269	8.38
R41N-R41H	121.692	8.588
H42N-H42H	122.373	8.526
E44N-E44H	121.645	8.911
A45N-A45H	125.627	8.463
A46N-A46H	123.763	8.463
S47N-S47H	115.297	8.362
A48N-A48H	126.258	8.383
A49N-A49H	125.241	8.367
G52N-G52H	109.492	8.648
A53N-A53H	123.727	8.249
S54N-S54H	114.829	8.429
L55N-L55H	124.537	8.226
L56N-L56H	120.463	7.986
L57N-L57H	120.643	7.868

9. Appendix

L58N-L58H	120.743	7.975
Q59N-Q59H	119.381	8.288
Q60N-Q60H	120.304	8.424
Q61N-Q61H	120.254	8.395
Q62N-Q62H	120.206	8.297
Q63N-Q63H	120.261	8.298
Q64N-Q64H	120.295	8.306
Q65N-Q65H	120.341	8.321
Q66N-Q66H	120.388	8.335
Q67N-Q67H	120.407	8.34
Q68N-Q68H	120.422	8.344
Q69N-Q69H	120.48	8.351
Q70N-Q70H	120.5	8.356
Q71N-Q71H	120.551	8.366
Q72N-Q72H	120.588	8.37
Q73N-Q73H	120.624	8.378
Q74N-Q74H	120.664	8.382
Q75N-Q75H	120.719	8.392
Q76N-Q76H	120.799	8.402
Q77N-Q77H	120.835	8.407
Q78N-Q78H	120.871	8.416
Q79N-Q79H	120.984	8.429
Q80N-Q80H	121.233	8.456
Q81N-Q81H	121.337	8.484
Q82N-Q82H	121.63	8.515
Q83N-Q83H	121.628	8.517
E84N-E84H	122.561	8.622
T85N-T85H	115.452	8.364

S86N-S86H	119.852	8.451
R88N-R88H	121.466	8.548
Q89N-Q89H	121.848	8.534
Q90N-Q90H	122.061	8.603
Q91N-Q91H	121.901	8.597
Q92N-Q92H	122.106	8.6
Q93N-Q93H	122.287	8.646
Q94N-Q94H	122.279	8.668
G95N-G95H	110.634	8.621
E96N-E96H	120.849	8.506
D97N-D97H	121.329	8.574
G98N-G98H	109.911	8.486
S99N-S99H	117.406	8.273
Q101N-Q101H	120.119	8.531
A102N-A102H	124.827	8.331
H103N-H103H	118.194	8.315
R104N-R104H	122.544	8.387
R105N-R105H	122.732	8.623
G106N-G106H	110.346	8.452
T108N-T108H	112.15	8.329
G109N-G109H	110.647	8.335
Y110N-Y110H	120.225	8.097
L111N-L111H	124.292	8.336
V112N-V112H	123.079	8.29
L113N-L113H	127.026	8.495
D114N-D114H	121.788	8.445
E115N-E115H	121.272	8.528
E116N-E116H	121.779	8.531

9. Appendix

Q117N-Q117H	121.418	8.456
Q118N-Q118H	122.898	8.594
S120N-S120H	117.029	8.626
Q121N-Q121H	122.98	8.54
Q123N-Q123H	121.343	8.746
S124N-S124H	117.546	8.52
A125N-A125H	126.333	8.536
L126N-L126H	120.77	8.21
E127N-E127H	121.417	8.33
C128N-C128H	120.621	8.367
H129N-H129H	122.79	8.681
E131N-E131H	121.55	9.103
R132N-R132H	122.386	8.457
G133N-G133H	109.967	8.538
C134N-C134H	119.045	8.344
V135N-V135H	124.739	8.45
E137N-E137H	123.058	8.637
G139N-G139H	109.684	8.681
A140N-A140H	123.79	8.109
A141N-A141H	123.61	8.429
V142N-V142H	120.232	8.217
A143N-A143H	128.166	8.489
A144N-A144H	123.646	8.423
S145N-S145H	115.363	8.381
K146N-K146H	123.41	8.445
G147N-G147H	109.683	8.434
L148N-L148H	123.049	8.254
Q150N-Q150H	120.935	8.616

Q151N-Q151H	122.533	8.581
L152N-L152H	126.183	8.579
A154N-A154H	126.21	8.508

9.5. pDEST-HisMBP vector

pDEST-HisMBP is a Gateway® destination vector for the production of recombinant proteins as fusions to the C-terminus of a dual His6-MBP tag in the cytoplasm of *E. coli* (340, 341). The vector is a derivative of pMal-C2 (new England Biolabs) with ColE1 replicon. The plasmid has antibiotic resistance for Ampicillin with a *tac* promoter. The expression of protein in this vector can be induced by IPTG (1mM).

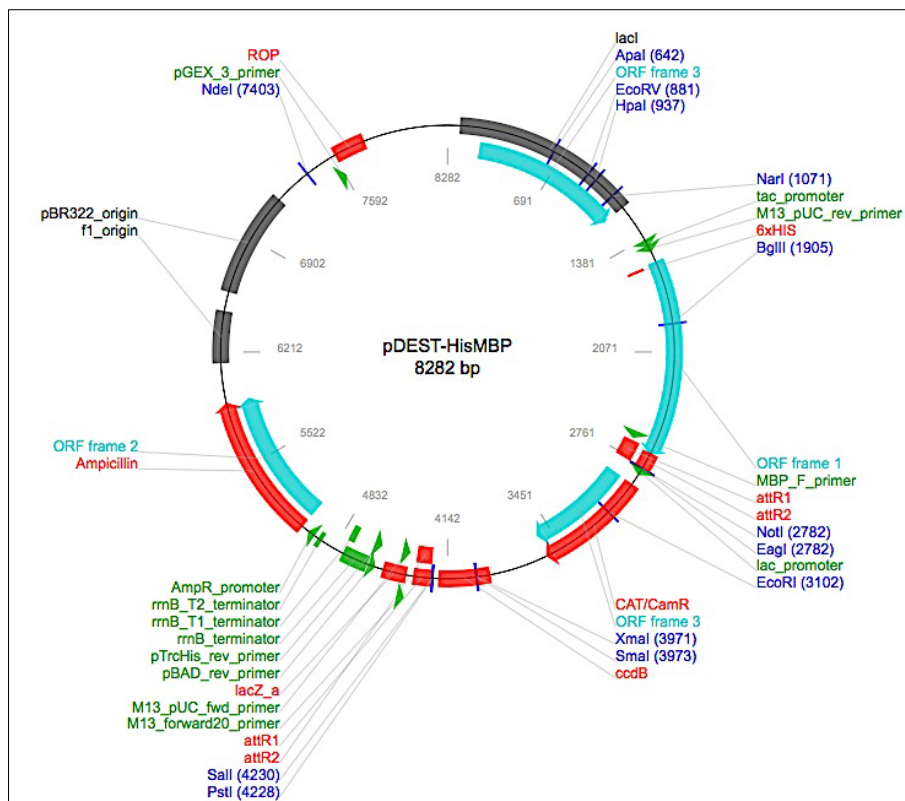


Figure 9-1: Scheme of pDEST-HisMBP vector.

9.6. Ni²⁺ affinity chromatography

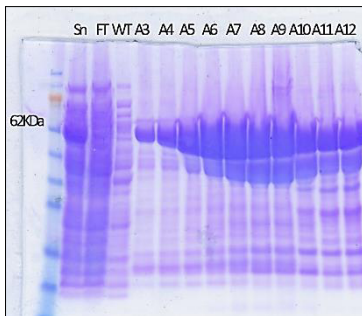
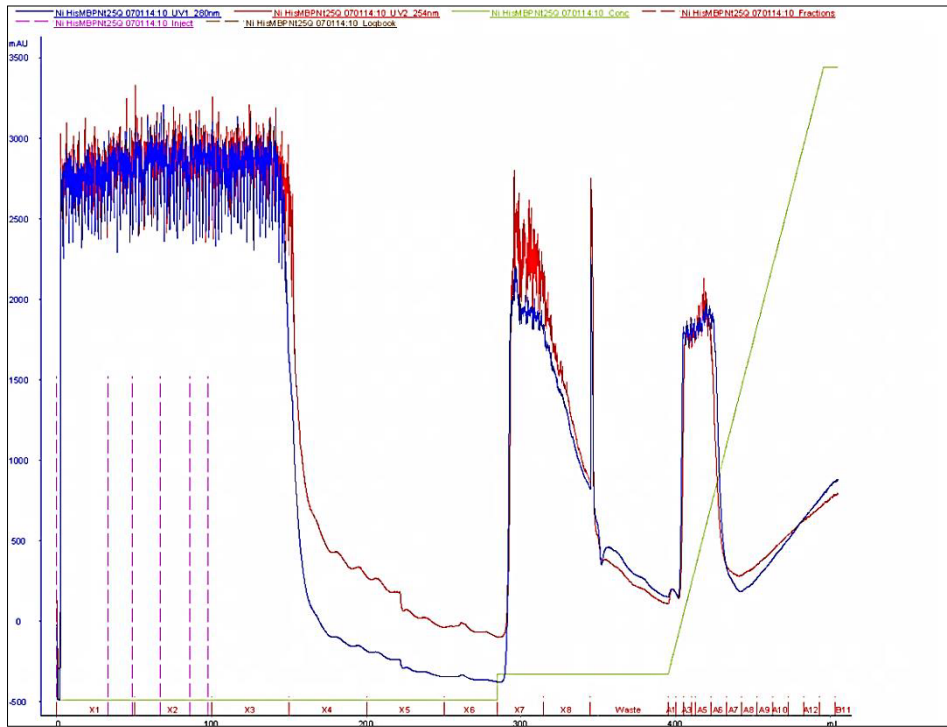


Figure 9-2: a) Ni²⁺ affinity chromatography and respective b) SDS PAGE gels of the more relevant fractions for HisMBP AR-25Q.

9.7. Size exclusion chromatography

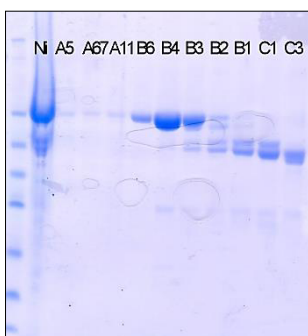
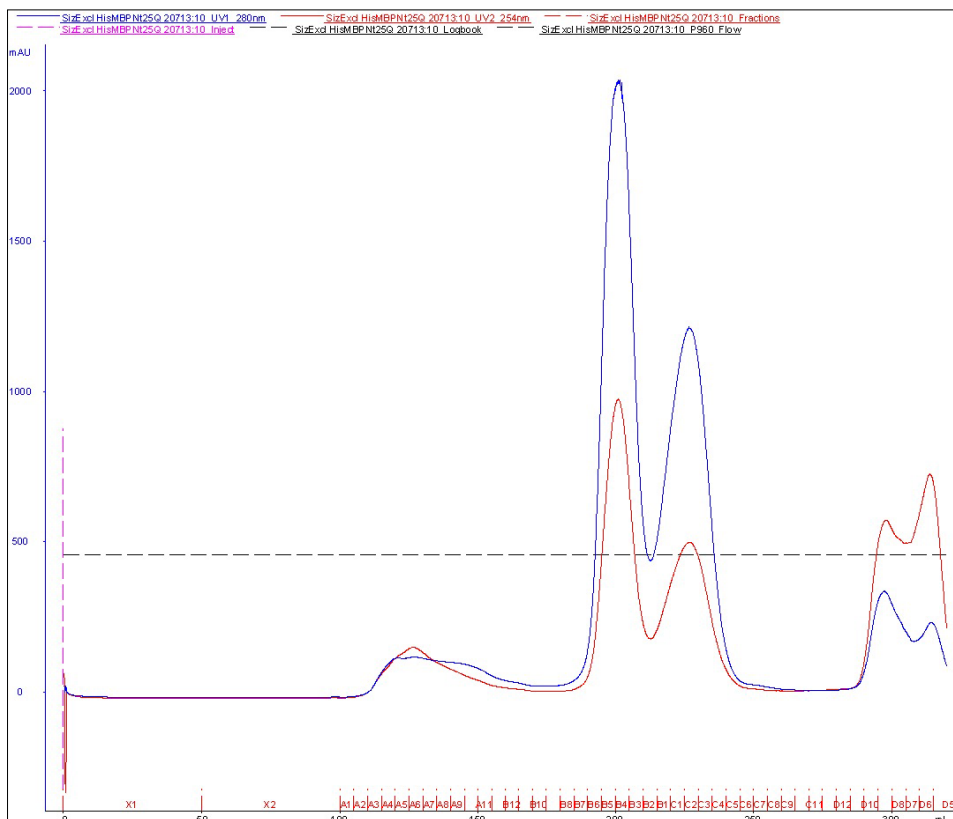


Figure 9-3: a) Size Exclusion chromatography and respective b) SDS PAGE gels of the more relevant fractions for HisMBP AR-25Q – HisMBP AR-25Q is the peak relative to fractions B6-B3.

9.8. Reverse Ni²⁺ chromatography

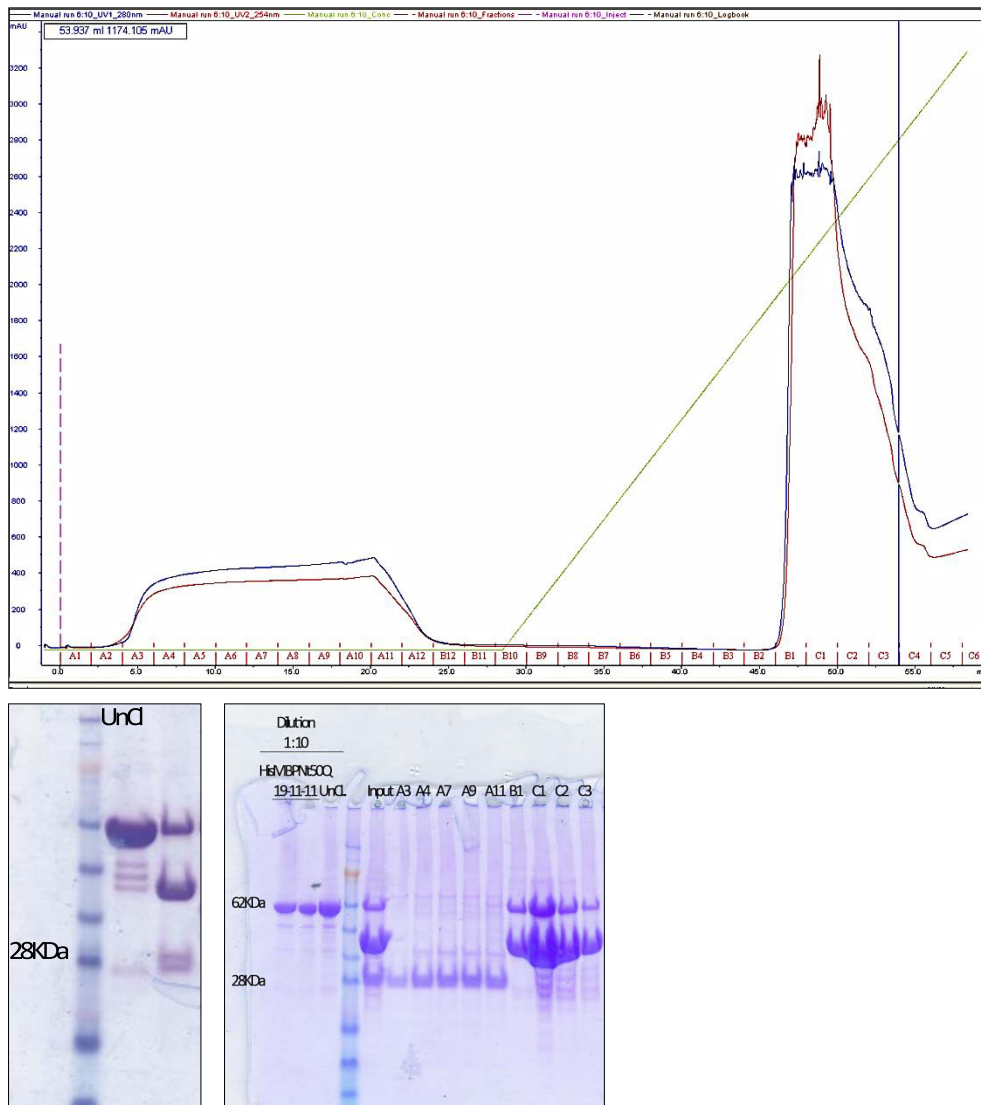


Figure 9-4: A representative reverse Ni²⁺ purification after TEV protease cleavage. a) chromatogram illustrates separated fractions of proteins, in this step, His-MBP and TEV proteins will be separated from AR-25Q. b) 2 SDS-PAGE gel examples of purified fractions, on left, His-MBP on the top and TEV protease are shown.

9.9. Mass spectrometric analysis of 4Q and 25Q

After each protein expression and purifications, the samples were sent for mass spectrometric (MS) analysis in mass spectroscopy facility in IRB, Barcelona.

Sample preparation for urea sample: Labeled ^{15}N AR-25Q and non-labeled protein prepared in 6M Urea: 10 μL of sample was desalted using a ZipTip[®] C4 and eluted with 10 μL of ACN/H₂O (70:30) with 1 % formic acid. The desalted sample was diluted 1/10 with 1 % formic acid aqueous solution and 10 μL were injected for MS analysis.

Sample preparation for non-urea sample: The sample was diluted with 1% formic acid aqueous solution to 5 μM and 10 μL (50 μmol) were injected for MS analysis.

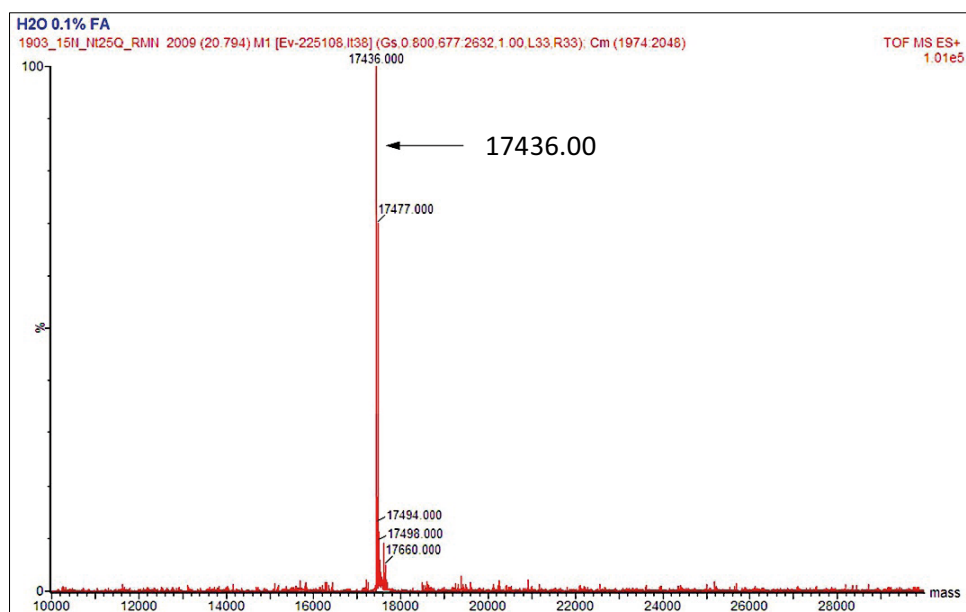


Figure 9-5: ^{15}N -25Q labeled sample for NMR experiments were prepared and the exact mass of the protein were identified before each experiment. **Isotopically labeled average molecular weight of AR-25Q is 17437. 1328.**

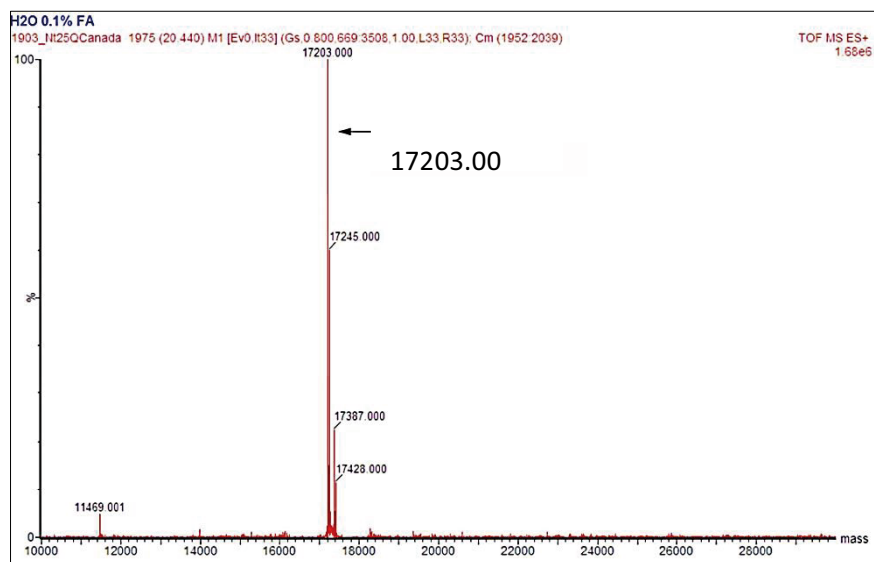


Figure 9-6: Non-labeled sample for NMR experiments was prepared and the exact mass of the protein was identified before each experiment. **The theoretical molecular weight of AR-25Q is 17203.91.**

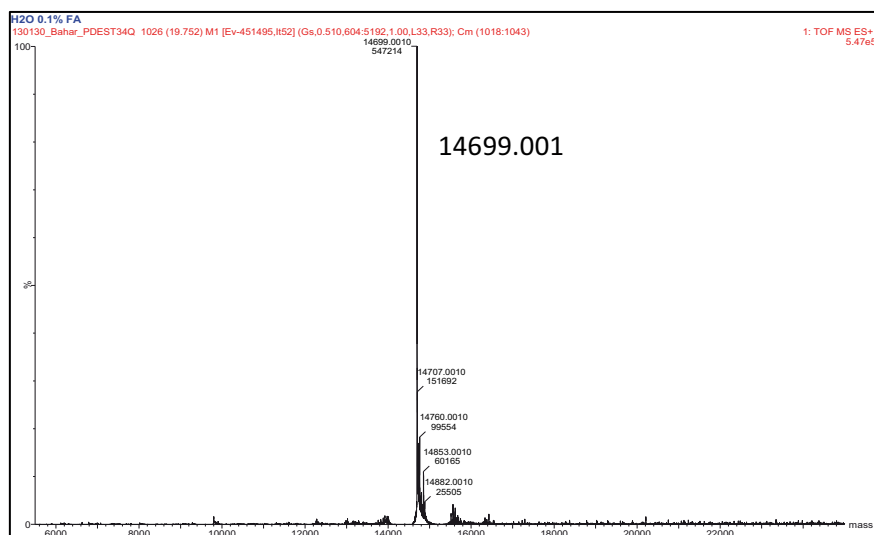


Figure 9-7: ^{15}N -4Q labeled sample for NMR experiments was prepared and the exact mass of the protein was identified before each experiment. **Isotopically labeled average molecular weight of AR-4Q is 14704.7051**

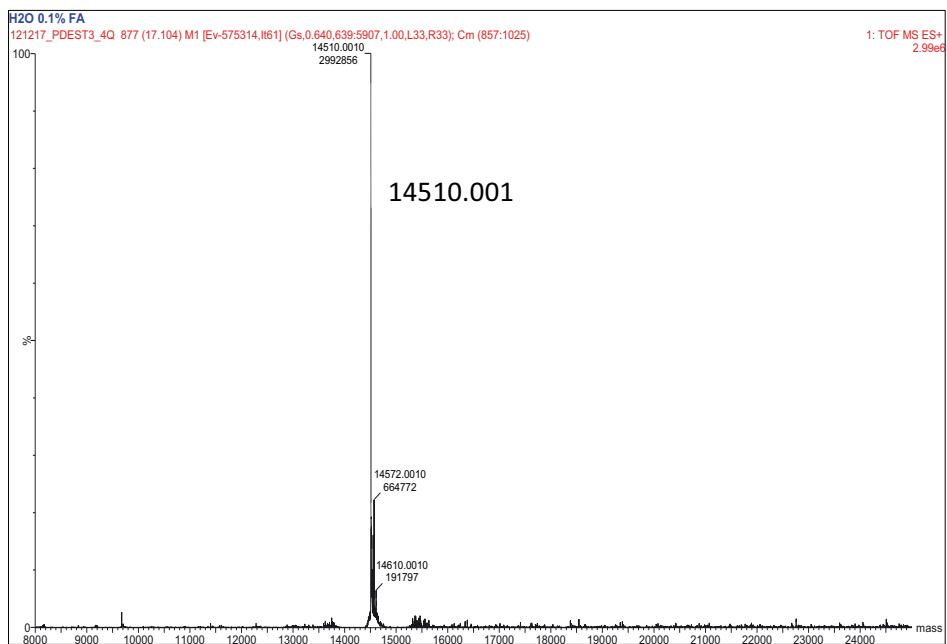


Figure 9-8: Non-labeled sample for NMR experiments was prepared and the exact mass of the protein was identified before each experiment. **The theoretical molecular weight of AR-25Q is 14513.1**

Table S1 Experimental parameters used for the acquisition of the NUS NMR experiments on the 4Q construct.

	Spectral widths and maximal evolution times		No. of scans	Inter-scan delays (s)	No. of complex points (aq)	No. of hypercomplex points	Duration of the experiment	Relative data points density (%)
	Indirect dimensions	Direct dimension						
4D HCBGACON	5000 Hz (¹ H α/β) 20.0 ms	12500 Hz (¹³ C α/β) 7.5 ms	8	0.9	512	850	1 day, 10 hours	0.07
4D HCBGACONCO	5000 Hz (¹ H α/β) 35.7 ms	12500 Hz (¹³ C α/β) 35.7 ms	16	0.9	512	850	3 days, 2 hours	0.11
4D (HCA)CON(CA)CON	2200 Hz (¹³ C) 24.1 ms	2600 Hz (¹⁵ N) 24.2 ms	16	0.9	512	930	3 days, 9 hours	0.34
4D (HN)CON(CA)CON	2200 Hz (¹³ C) 24.1 ms	2600 Hz (¹⁵ N) 24.2 ms	32	0.5	512	910	4 days, 11 hours	0.33
3D TROSY HNCO		2700 Hz (¹³ C) 20.7 ms	8	1.2	1024	560	7 hours	20.00
4D TROSY (H)NCO(CA)NNH	2300 Hz (¹⁵ N) 20.4 ms	2700 Hz (¹³ C) 23.9 ms	8	1.2	1024	2660	2 days, 20 hours	1.60
4D TROSY HN(CO)CA)NNH	1500 Hz (¹ H) 20.0 ms	2300 Hz (¹⁵ N) 23.9 ms	8	1.2	1024	1450	1 day, 13 hours	1.60

Table S2 Experimental parameters used for the acquisition of the ^{15}N relaxation NMR experiments on the 4Q construct.

	Spectral widths and maximal evolution times		No. of scans	Inter-scan delays (s)
2D ^1H - ^{15}N HSQC	2100 Hz (^{15}N) 24.0 ms	8400 Hz (^1H) 61.1 ms	2	1.0
2D CON-IPAP	2800 Hz (^{15}N) 90.0 ms	5300 Hz ($^{13}\text{C}'$) 96.8 ms	8	2.5
The spectra were acquired at the following temperatures: 278, 283, 288, 293, 298, 303 and 308 K.				

Table S3 Experimental parameters used for the acquisition of the time dependence NMR experiments on the 4Q construct.

	Spectral widths and maximal evolution times		No. of scans	Inter-scan delays (s)
^{15}N R_1	1600 Hz (^{15}N) 156.8 ms	10500 Hz (^1H) 97.6 ms	8	3.0
^{15}N R_2	1600 Hz (^{15}N) 156.8 ms	10500 Hz (^1H) 97.6 ms	8	3.0
Steady-state heteronuclear $^{15}\text{N}\{^1\text{H}\}$ NOEs	1600 Hz (^{15}N) 156.8 ms	10500 Hz (^1H) 97.6 ms	64	6.0
For the determination of R_1 , 10 experiments were acquired changing the variable delay from 15 to 995 ms. For the determination of R_2 , 10 experiments were acquired changing the variable delay from 30 to 565 ms.				

Table S4 Experimental parameters used for the acquisition of the NUS NMR experiments on the 25Q construct.

	Spectral widths and maximal evolution times		Direct dimension	No. of scans	Inter-scan delays (s)	No. of complex points (aq)	No. of hypercomplex points	Duration of the experiment	Relative data points density (%)
	Indirect dimensions								
4D HCBGACON	5000 Hz (¹ H ^{α/β}) 20.0 ms	12500 Hz (¹³ C ^{α/β}) 7.5 ms	2600 Hz (¹⁵ N) 50.0 ms	8	0.9	512	850	1 day, 10 hours	0.07
4D (HN)CON(CA)CON	2200 Hz (¹³ C) 24.1 ms	2600 Hz (¹⁵ N) 24.2 ms	2600 Hz (¹⁵ N) 30.0 ms	32	0.5	512	910	4 days, 11 hours	0.33
3D HNCO		2000 Hz (¹³ C) 35.5 ms	2000 Hz (¹⁵ N) 24.0 ms	16	1.0	1024	1130	1 day	33.30
3D TROSY HN(CA)CO		1800 Hz (¹³ C) 25.0 ms	2400 Hz (¹⁵ N) 20.8 ms	16	1.2	1024	580	14 hours	26.00
4D TROSY HN(COCA)NNH	1500 Hz (¹ H) 20.0 ms	2400 Hz (¹⁵ N) 22.9 ms	2400 Hz (¹⁵ N) 22.9 ms	8	1.2	1024	1720	1 day, 20 hours	1.90

Table S5 Experimental parameters used for the acquisition of the ^{15}N relaxation NMR experiments on the 0.5 mM 25Q construct.

	Spectral widths and maximal evolution times		No. of scans	Inter-scan delays (s)
^{15}N R_1	1600 Hz (^{15}N) 156.8 ms	10500 Hz (^1H) 97.6 ms	8	3.0
^{15}N R_2	1600 Hz (^{15}N) 156.8 ms	10500 Hz (^1H) 97.6 ms	8	3.0
Steady-state heteronuclear $^{15}\text{N}\{^{15}\text{H}\}$ NOEs	1600 Hz (^{15}N) 177.7 ms	10500 Hz (^1H) 97.6 ms	64	6.0
<p>For the determination of R_1, 10 experiments were acquired changing the variable delay from 15 to 995 ms. For the determination of R_2, 10 experiments were acquired changing the variable delay from 30 to 315 ms.</p> <p>R_2 values were determined also for 0.25 and 0.125 mM samples by acquiring, respectively, 7 (16 scans) and 9 (24 scans) experiments changing the variable delay from 30 to 315 ms.</p>				

Table S6 list of proteins with high abundance isolated from total tissue homogenates of SBMA transgenic mice.

	spinal cord young	spinal cord old	spinal cord old	spinal cord old	muscle young	muscle old	muscle very old			
	TP_XS_3	TP_XS_1	TP_XS_2	TP_XS_4	TP_XS_5	TP_XS_7	TP_XS_6			
	10812	10810	10811	10816	10813	10815	10814			
UniProt ID								TOTAL peptides	Biological function	
									Protein type	
sp A2ASS6-2 TTTN_MOUSE	5					95	130	230	adult heart development cardiac myofibril assembly forward locomotion heart morphogenesis in utero embryonic development muscle contraction peptidyl-tyrosine phosphorylation regulation of relaxation of cardiac muscle	Kinase Serine/threonine- protein kinase Transferase
sp A2ASS6 TTTN_MOUSE	11					37	53	101	adult heart development cardiac myofibril assembly forward locomotion heart morphogenesis in utero embryonic development muscle contraction peptidyl-tyrosine phosphorylation regulation of relaxation of cardiac muscle	Kinase Serine/threonine- protein kinase Transferase
sp O08553 DPYL2_MOUSE	5	29	31	49	28	1	3	146	cytoskeleton organization endocytosis olfactory bulb development positive regulation of glutamate secretion pyrimidine nucleobase catabolic process regulation of neuron differentiation response to amphetamine response to cocaine response to drug spinal cord development synaptic vesicle transport	Developmental protein

SP O08599 2 STXB1_MOUSE	10	25	19	12	1	67	axon target recognition long term synaptic depression negative regulation of neuron apoptotic process negative regulation of synaptic transmission, GABAergic neuromuscular synaptic transmission neurotransmitter secretion positive regulation of calcium ion-dependent exocytosis protein transport synaptic vesicle maturation vesicle docking involved in exocytosis		
SP O08599 STXB1_MOUSE	1	11	25	12	24	1	1	75 DNA unwinding involved in DNA replication apoptotic process cell differentiation cell proliferation negative regulation of transcription, DNA-dependent transcription, DNA-dependent	
SP O35295 PUBB_MOUSE	6	13	15	12	21	1	7	75 actin cytoskeleton organization cortical actin cytoskeleton organization regulation of cell shape	Repressor
SP O70318 E41L2_MOUSE		10	12	16	10		2	50 RNA splicing RNA transport mRNA processing negative regulation of mRNA splicing, via spliceosome	
SP O08569 RNA2_MOUSE	8	8	5	5	8	4	8	56 RNA splicing RNA transport mRNA processing negative regulation of mRNA splicing, via spliceosome	Ribonucleoprotein

SP P08553 NFM_MOUSE	3	51	44	59	33	11	201	axon cargo transport intermediate filament bundle assembly microtubule cytoskeleton organization neurofilament cytoskeleton organization regulation of axon diameter	
SP P09411 PGK1_MOUSE	16	3	4	13	3	6	8	glycolysis	Kinase Transferase
SP P10922 H10_MOUSE	5	11	10	6	2	4	13	nucleosome assembly	
SP P11499 HS90B_MOUSE	5	8	6	6	7	1	4	cellular response to interleukin-4 negative regulation of proteasomal ubiquitin-dependent protein catabolic process placenta development positive regulation of nitric oxide biosynthetic process protein folding regulation of interferon-gamma- mediated signaling pathway regulation of type I interferon- mediated signaling pathway response to stress	Chaperone
SP P13542 MYH8_MOUSE	16		5			33	45	ATP binding motor activity	Motor protein Muscle protein Myosin

sp P14873 MAP1B_MOUSE		23	21	35	25				104	axon extension dendrite development establishment of monopolar cell polarity microtubule bundle formation mitochondrion transport along microtubule negative regulation of intracellular transport positive regulation of axon extension	
sp P16330-2 CNS7_MOUSE	1	17	16	35	32		7	108	adult locomotory behavior aging axonogenesis cyclic nucleotide catabolic process forebrain development microtubule cytoskeleton organization regulation of mitochondrial membrane permeability response to lipopolysaccharide response to toxic substance	Hydrolase	
sp P16330 CNS7_MOUSE	4	15	26	13	15			73	adult locomotory behavior aging axonogenesis cyclic nucleotide catabolic process forebrain development microtubule cytoskeleton organization regulation of mitochondrial membrane permeability response to lipopolysaccharide response to toxic substance	Hydrolase	
sp P16558 GAP_MOUSE	51	30	31	28	24	39	40	244	glycolysis microtubule cytoskeleton organization negative regulation of translation neuron apoptotic process peptidyl-L-cysteine S-trans-nitrosylation protein stabilization	Oxidoreductase Transferase	

SP P17156 HSP72_MOUSE	6	9	7	7	9	6	7	71	male meiosis I positive regulation of cyclin-dependent protein serine/threonine kinase activity involved in G2/M transition of mitotic cell cycle response to stress spermatid development synaptonemal complex disassembly	Chaperone
SP P17182 ENOA_MOUSE	10	23	20	18	25	17	16	130	glycolysis cellular response to cell-matrix adhesion cellular response to epidermal growth factor stimulus cellular response to glucocorticoid stimulus cellular response to insulin stimulus cellular response to lipopolysaccharide common myeloid progenitor cell proliferation glutathione metabolic process negative regulation of ERK1 and ERK2 cascade negative regulation of I-kappaB kinase/NF-kappaB cascade negative regulation of JUN kinase activity negative regulation of acute inflammatory response negative regulation of apoptotic process negative regulation of extrinsic apoptotic signaling pathway negative regulation of fibroblast proliferation negative regulation of interleukin-1 beta production	Lyase
SP P19157 GSTP1_MOUSE	12	5	10	12	9	14	17	79		Transferase

sp P19246 NFH_MOUSE	1	67	64	78	36	7	253	intermediate filament bundle assembly microtubule cytoskeleton organization neurofilament cytoskeleton organization peripheral nervous system neuron axonogenesis	
sp P20029 GRP78_MOUSE	9	9	10	8	10	2	9	70 ER overload response activation of signaling protein activity involved in unfolded protein response cellular response to glucose starvation cellular response to interleukin-4 cerebellar Purkinje cell layer development cerebellum structural organization negative regulation of apoptotic process negative regulation of transforming growth factor beta receptor signaling pathway positive regulation of embryonic development positive regulation of protein ubiquitination proteolysis involved in cellular protein catabolic process	
sp P20152 VIME_MOUSE	25	32	36	44	34	8	20	209 Bergmann glial cell differentiation astrocyte development intermediate filament organization lens fiber cell development negative regulation of neuron projection development	
sp P21550 ENOB_MOUSE	61	1		6	9	56	63	196 aging glycolysis response to drug skeletal muscle tissue regeneration	Lyase
sp P22752 H2A1_MOUSE	7	4	5	11	4	3	8	61 nucleosome assembly	

sp P23927 CRVAB_MOUSE	8				1	44	24	77	apoptotic process involved in morphogenesis cellular response to gamma radiation glucose metabolic process lens development in camera-type eye microtubule polymerization or depolymerization muscle organ development negative regulation of apoptotic process negative regulation of cell growth negative regulation of cysteine-type endopeptidase activity involved in apoptotic process negative regulation of gene expression negative regulation of intracellular transport negative regulation of reactive oxygen species metabolic process protein folding protein homooligomerization response to estradiol stimulus response to hydrogen peroxide response to hypoxia stress-activated MAPK cascade tubulin complex assembly	Chaperone Eye lens protein
sp P26443 DHE3_MOUSE		36	30	37	17			120	glutamine metabolic process long-term memory positive regulation of insulin secretion response to aluminum ion tricarboyllic acid metabolic process	Oxidoreductase
sp P31001 DESM_MOUSE	26					95	88	209	muscle organ development	Muscle protein
sp P35486 ODPA_MOUSE	44	38	45	47	43	11	36	264	acetyl-CoA biosynthetic process from pyruvate glycolysis tricarboyllic acid cycle	Oxidoreductase

sp P42669 PURA_MOUSE	3	23	15	16	21	2	5	85	DNA unwinding involved in DNA replication apoptotic process cell differentiation cell proliferation mitotic cell cycle checkpoint negative regulation of transcription, DNA-dependent nervous system development positive regulation of cell proliferation transcription, DNA-dependent	Activator
sp P45591 COF2_MOUSE	16	13	14	14	15	15	19	110	actin filament depolymerization positive regulation of receptor recycling potassium ion transport protein transport	
sp P46460 NSF_MOUSE		10	34	24	27			95	cell differentiation nervous system development neurofilament cytoskeleton organization	Hydrolase
sp P46660 ANX_MOUSE	6	73	62	89	44		1	275	intracellular protein transport	Developmental protein
sp P47708 RP3A_MOUSE		8	8	11	27			54	fructose 6-phosphate metabolic process glucose homeostasis glycolysis muscle cell cellular homeostasis positive regulation of insulin secretion	Kinase Transferase
sp P47857 K6PF_MOUSE	25	15	14	10	10	20	22	116		

sp P47962 RL5_MOUSE	6	14	10	6	12	2	7	57	RNA processing ribosomal large subunit biogenesis translation	Ribonucleoprotein Ribosomal protein
sp P48036 ANXA5_MOUSE	14	25	27	27	17	23	25	160	blood coagulation negative regulation of coagulation response to organic substance	
sp P48758 CBRL_MOUSE	5	9	12	6	5	10	11	58	drug metabolic process vitamin K metabolic process	Oxidoreductase
sp P48962 ADTL_MOUSE	19	7	14	8	14	5	12	82	apoptotic mitochondrial changes negative regulation of necrotic cell death transmembrane transport	
sp P50518 VATE1_MOUSE	5	28	39	36	36	4	7	155	ATP hydrolysis coupled proton transport	Hydrolase
sp P51881 ADT2_MOUSE	4	5	11	8	13	2	4	53	chromosome segregation negative regulation of mitochondrial outer membrane permeabilization positive regulation of cell proliferation transmembrane transport	
sp P52480-2 KPYM_MOUSE	12	15	10	40	8	1	6	99	glycolysis	Kinase Transferase
sp P52480 KPYM_MOUSE	20	3	10	60	17	11	9	147	glycolysis	Kinase Transferase

sp P62702 RS4X_MOUSE	9	11	16	18	23	3	3	3	85	negative regulation of megakaryocyte differentiation nucleosome assembly	Ribonucleoprotein Ribosomal protein
sp P62806 H4_MOUSE	10	7	3	10	2	11	17	73	ATP hydrolysis coupled proton transport		
sp P62814 VATB2_MOUSE		14	11	14	12		1	52	ATP hydrolysis coupled proton transport ATP metabolic process	Hydrolase	
sp P56480 ATPB_MOUSE	36	19	15	30	25	5	19	166	ATP hydrolysis coupled proton transport ATP synthesis coupled proton transport angiogenesis lipid metabolic process negative regulation of cell adhesion involved in substrate-bound cell migration regulation of intracellular pH	Hydrolase	
sp P57746 VATD_MOUSE	8	36	34	41	30	2	11	164	cilium assembly protein localization to cilium proton transport		
sp P62424 RS8_MOUSE	9	12	13	10	11	3	13	71	translation	Ribonucleoprotein Ribosomal protein	
sp P62301 RS13_MOUSE	27	22	25	26	34	10	18	162	negative regulation of RNA splicing translation	Ribonucleoprotein Ribosomal protein	

sp P62827 RAN_MOUSE	6	6	7	10	7	4	5	50	actin cytoskeleton organization cell division cellular protein complex localization mitosis positive regulation of protein binding protein import into nucleus small GTPase mediated signal transduction	
sp P62900 RL31_MOUSE	8	17	15	22	19	5	7	94	translation	Ribonucleoprotein Ribosomal protein
sp P62908 RS3_MOUSE	5	11	11	9	11	1	3	51	translation	Ribonucleoprotein Ribosomal protein
sp P62918 RL9_MOUSE	13	20	22	20	22	1	13	114	translation actin cytoskeleton organization cellular response to growth factor stimulus neural tube closure positive regulation of DNA metabolic process positive regulation of stress fiber assembly positive regulation of transcription from RNA polymerase II promoter positive regulation of viral transcription regulation of transcription from RNA polymerase II promoter sequestering of actin monomers	Ribonucleoprotein Ribosomal protein
sp P62962 PROF1_MOUSE	7	10	8	11	7	11	11	66	RNA splicing chaperone mediated protein folding requiring cofactor mRNA processing negative regulation of transcription, DNA-dependent regulation of cell cycle response to stress transcription, DNA-dependent	Chaperone Repressor
sp P63017 HSP7C_MOUSE	22	23	30	33	29	15	18	183		

sfp63038-2 CH60_MOUSE	8	9	12	8	13	9	8	69	<p>B cell cytokine production B cell proliferation MyD88-dependent toll-like receptor signaling pathway T cell activation activation of cysteine-type endopeptidase activity involved in apoptotic process isotype switching to IgG isotypes negative regulation of apoptotic process positive regulation of T cell activation positive regulation of T cell mediated immune response to tumor cell positive regulation of apoptotic process positive regulation of interferon-alpha production positive regulation of interferon-gamma production positive regulation of interleukin-10 production positive regulation of interleukin-12 production positive regulation of interleukin-6 production positive regulation of macrophage activation</p>	Chaperone
-----------------------	---	---	----	---	----	---	---	----	--	-----------

sp P63039 CH60_MOUSE	5	8	8	7	7	3	7	59	B cell cytokine production B cell proliferation MyD88-dependent toll-like receptor signaling pathway T cell activation activation of cysteine-type endopeptidase activity involved in apoptotic process isotype switching to IgG isotypes negative regulation of apoptotic process positive regulation of T cell activation positive regulation of T cell mediated immune response to tumor cell positive regulation of apoptotic process positive regulation of interferon-alpha production positive regulation of interferon-gamma production positive regulation of interleukin-10 production positive regulation of interleukin-12 production positive regulation of interleukin-6 production positive regulation of macrophage activation	Chaperone
sp P63276 RS17_MOUSE	14	20	17	19	26	2	7	105	ribosomal small subunit assembly translational elongation	Ribonucleoprotein Ribosomal protein
sp P67984 RL22_MOUSE	7	10	13	15	13	2	7	67	alpha-beta T cell differentiation translational	Ribonucleoprotein Ribosomal protein
sp P70404 IDHG1_MOUSE	22	14	17	19	20	12	16	120	isocitrate metabolic process tricarboxylic acid cycle	Oxidoreductase
sp Q04447 KCRB_MOUSE	3	22	24	29	25	2	5	110	brain development cellular chloride ion homeostasis	Kinase Transferase
sp Q55X39 MYH4_MOUSE	27		3			43	40	113	muscle contraction response to activity	Motor protein Muscle protein Myosin
sp Q55X40 MYH1_MOUSE	11					43	37	91	ATP binding motor activity	Motor protein Muscle protein Myosin

SP Q614Z5 HCDH_MOUSE	20	11	8	9	10	20	18	96	fatty acid beta-oxidation negative regulation of insulin secretion response to activity response to drug response to insulin stimulus	Oxidoreductase
SP Q62344 MYOM1_MOUSE	9					25	19	53	muscle contraction muscle contraction	Muscle protein
SP Q624Z5 NDUA4_MOUSE	19	20	16	21	18	7	14	115	muscle contraction	
SP Q647Z7 VINC_MOUSE	23	9	9	7	8	26	34	117	adherens junction assembly apical junction assembly cell adhesion epithelial cell-cell adhesion lamellipodium assembly morphogenesis of an epithelium protein localization to cell surface regulation of cell migration	
SP Q6P8J7 KCGR_MOUSE	34				2	3	31	94	phosphocreatine metabolic process	Kinase Transferase
SP Q7TME3 NDUAC_MOUSE	13	10	12	11	14	4	8	72	response to oxidative stress	
SP Q7TQ48-2 SRCA_MOUSE	26					6	19	51	GTP catabolic process	
SP Q8BMF4 ODP2_MOUSE	23	18	18	28	30	5	7	129	acetyl-CoA biosynthetic process from pyruvate glucose metabolic process tricarboxylic acid cycle	Acyltransferase Transferase
SP Q8BMS1 EGHA_MOUSE	32	37	44	15	39	31	57	255	fatty acid beta-oxidation response to insulin stimulus	Lyase Oxidoreductase
SP Q8BW11 THIM_MOUSE	17	9	6	9	7	7	14	69	fatty acid metabolic process negative regulation of mitochondrial outer membrane permeabilization	Acyltransferase Transferase

sp Q8QZT1 THIL_MOUSE	7	16	15	19	17	2	4	80	adipose tissue development brain development liver development metanephric proximal convoluted tubule development protein homooligomerization response to hormone stimulus response to organic cyclic compound response to starvation	Acyltransferase Transferase
sp Q8R164 BPHL_MOUSE	8	10	10	13	14	13	16	84	hydrolyase activity	Hydrolase
sp Q99YU0 EGHE_MOUSE	26	16	16	18	20	16	24	138	fatty acid beta-oxidation	Acyltransferase Transferase
sp Q99K10 ACON_MOUSE	29	21	18	49	24	21	33	195	citrate metabolic process isocitrate metabolic process tricarboxylic acid cycle	Lyase
sp Q991C3 NDUAA_MOUSE	7	10	13	9	12	1	7	59	nucleobase-containing compound metabolic process oxidation-reduction process response to drug	
sp Q9CF08 ATP5L_MOUSE	7	4	6	10	14	3	8	52	ATP synthesis coupled proton transport	
sp Q9CQ69 QCR8_MOUSE	10	8	7	11	11	6	9	62	cerebellar Purkinje cell layer development hippocampus development hypothalamus development midbrain development pons development pyramidal neuron development subthalamus development thalamus development	
sp Q9CQC7 NDUB4_MOUSE	24	19	22	21	27	14	18	145	response to oxidative stress	

sp Q9CQJ8 NDUB9_MOUSE	15	13	15	17	15	6	11	92	oxidation-reduction process	
sp Q9CQZ7 AT5F1_MOUSE	13	5	12	11	11	3	5	60	ATP synthesis coupled proton transport	
sp Q9CQZ5 NDUA6_MOUSE	13	11	10	19	15	3	5	76	oxidation-reduction process response to oxidative stress	
sp Q9D1B9 RMZ8_MOUSE	8	7	10	11	10	2	5	53	translation	Ribonucleoprotein Ribosomal protein
sp Q9DBE6 RL4_MOUSE	12	21	20	14	15	3	9	99	translation	Ribonucleoprotein Ribosomal protein
sp Q9DB20 ATPO_MOUSE	20	15	16	21	22	9	13	116	ATP synthesis coupled proton transport	
sp Q9DB77 QCR2_MOUSE	37	32	26	34	38	12	30	209	oxidation-reduction process proteolysis	
sp Q9DC69 NDUA9_MOUSE	22		13	10	13	3	4	65	oxidation-reduction process	
sp Q9DC70 NDUS7_MOUSE	14	10	8	13	14	3	8	70	mitochondrial respiratory chain complex I assembly	Oxidoreductase
sp Q9DCW4 ETFB_MOUSE	24	16	17	17	13	25	26	138	oxidation-reduction process	
sp Q9J181 RL38_MOUSE	5	11	10	8	15			51	90S preribosome assembly axial mesoderm development middle ear morphogenesis ossification regulation of translation sensory perception of sound skeletal system development translation	Ribonucleoprotein Ribosomal protein

sp Q91KB3-Z YBOX3_MOUSE		18	4	8	10	14	8	18	80	cellular hypertonic response cellular response to tumor necrosis factor fertilization in utero embryonic development male gonad development negative regulation of apoptotic process negative regulation of intrinsic apoptotic signaling pathway in response to osmotic stress negative regulation of necrotic cell death negative regulation of skeletal muscle tissue development positive regulation of organ growth regulation of transcription, DNA-dependent spermatogenesis transcription, DNA-dependent	Repressor
sp Q9R0Y5-Z KAD1_MOUSE		26	1	3	2	13	25	16	86	ATP metabolic process cell cycle arrest nucleoside triphosphate biosynthetic process	Kinase Transferase
sp Q9R0Y5 KAD1_MOUSE		41	13	11	12	9	49	51	186	ATP metabolic process cell cycle arrest nucleoside triphosphate biosynthetic process	Kinase Transferase
sp Q9WTF7 KAD3_MOUSE		11	12	13	10	12	7	13	78	AMP metabolic process GTP metabolic process UTP metabolic process	Kinase Transferase
sp Q9WUM5 SUGA_MOUSE		47	38	34	32	38	32	41	263	succinate metabolic process succinyl-CoA metabolic process tricarboxylic acid cycle	Ligase
sp Q9Z0U1 Z0Z_MOUSE			22	24	12	25			83		
sp Q9Z1P6 NDUA7_MOUSE		15	11	8	8	16	3	9	70	ATP synthesis coupled electron transport	
sp Q9Z219 SUCB1_MOUSE		19	15	13	15	13	8	14	97	tricarboxylic acid cycle	Ligase

sp P01942 HBA_MOUSE	18	19	21	29	19	23	33	162	response to stilbenoid Bergmann glial cell differentiation astrocyte development extracellular matrix organization intermediate filament organization long-term synaptic potentiation negative regulation of neuron projection development neuron projection regeneration positive regulation of Schwann cell proliferation regulation of neurotransmitter uptake	
sp P03995-2 GFAP_MOUSE	1	16	39	39	18	1	1	115	regulation of neurotransmitter uptake response to wounding Bergmann glial cell differentiation astrocyte development extracellular matrix organization intermediate filament organization long-term synaptic potentiation negative regulation of neuron projection development neuron projection regeneration positive regulation of Schwann cell proliferation regulation of neurotransmitter uptake response to wounding	
sp P03995 GFAP_MOUSE	5	68	47	97	53	2	3	275	aspartate biosynthetic process aspartate catabolic process fatty acid transport glutamate catabolic process to 2-oxoglutarate glutamate catabolic process to aspartate oxaloacetate metabolic process response to ethanol	Aminotransferase Transferase
sp P05202 AATM_MOUSE	21			15	17			53		

† B0VZRH4 B0VZRH4_MOUSE	2	17	19	27	25		2	92	hydrogen-exporting ATPase activity, phosphorylative mechanism	
† B0VZNS B0VZNS_MOUSE	6	9	13	12	9	8	20	83	angiogenesis cellular response to acid collagen fibril organization fibrinolysis positive regulation of binding	
† B1AR69 B1AR69_MOUSE	11		3			33	37	84	cellular response to starvation	Motor protein Myosin
† B1AR43 B1AR43_MOUSE	9	8	10	12	10	3	4	56	RNA processing ribosomal large subunit biogenesis translation	
† B1ARR7 B1ARR7_MOUSE	3	11	10	10	11	2	4	54	glycolysis	
† B1ATY1 B1ATY1_MOUSE	15	17	17	13	19	11	13	129	sarcomere organization	
† B2RST7 B2RST7_MOUSE	26				1	14	24	65	skeletal muscle cell differentiation	
† B9EHJ3 B9EHJ3_MOUSE		15	10	7	20			52	blastocyst formation	
† D3YTQ9 D3YTQ9_MOUSE	8	7	5	11	17	2	5	57	RNA processing ribosomal small subunit biogenesis ribosomal small subunit export from nucleus translation	Ribonucleoprotein Ribosomal protein
† D3YU93 D3YU93_MOUSE	9	13	12	12	13	6	7	72	ribosome biogenesis	
† D3YUG3 D3YUG3_MOUSE	12	17	16	16	24	4	9	98	translation	
† D3YV43 D3YV43_MOUSE	7	6	6	10	13	1	5	50	disaperone mediated protein folding requiring cofactor regulation of cell cycle	Ribonucleoprotein Ribosomal protein
† D3YW43 D3YW43_MOUSE	10	16	14	18	17	8	12	111		

TI D3Z2AN9ID3Z21N9_MOUSE	7	17	16	16	19	4	8	87	translational B cell cytokine production B cell proliferation MyD88-dependent toll-like receptor signaling pathway T cell activation activation of cysteine-type endopeptidase activity involved in apoptotic process isotype switching to IgG isotypes negative regulation of apoptotic process positive regulation of T cell mediated immune response to tumor cell positive regulation of apoptotic process positive regulation of interferon-alpha production positive regulation of interferon-gamma production positive regulation of interferon-10 production positive regulation of interleukin-12 production positive regulation of interleukin-6 production positive regulation of macrophage activation protein refolding	Chaperone
TI D3Z2ZF1D3Z2FZ_MOUSE	6	14	15	14	13	7	14	85	cell differentiation translational	Ribonucleoprotein Ribosomal protein
TI D3Z6C3ID3Z6C3_MOUSE	25	20	27	30	34	7	17	160	ATP hydrolysis coupled proton transport ATP synthesis coupled proton transport embryo development lipid metabolic process negative regulation of endothelial cell proliferation	
TI D3Z6F9ID3Z6F5_MOUSE	101	134	140	106	123	61	87	761	microtubule cytoskeleton organization	
TI D3Z6W1ID3Z6W1_MOUSE		17	19	34	16			86	glycogen anabolic process	Glycosyltransferase Transferase
TI E9PUM3IE9PUM3_MOUSE	69		4	4	3	26	39	145		

U E9PV44 E9PV44_MOUSE	3	10	12	7	13	2	5	52	erythrocyte differentiation heme biosynthetic process negative regulation of catalytic activity negative regulation of endothelial cell proliferation negative regulation of nucleotide metabolic process positive regulation of mitochondrial outer membrane permeabilization reactive oxygen species metabolic process	Glycoyltransferase Transferase
U E9PV44 E9PV44_MOUSE	3	10	12	7	13	2	5	52	actin cytoskeleton organization	
U E9PZF4 E9PZF4_MOUSE	2					67	65	134	microtubule-based movement	
U E9QJ15 E9QJ15_MOUSE		19	26	22	23		1	91	binding of sperm to zona pellucida protein folding	Measur protein
U E9Q133 E9Q133_MOUSE	8	13	14	13	10		5	63	cell cycle protein heterooligomerization	Chaperone
U E9Q1G8 E9Q1G8_MOUSE	3	22	28	31	28		5	117	axoanogenesis	
U E9Q5F4 E9Q5F4_MOUSE	1	5	5	6	6		2	53	membrane organization myelination response to toxic substance	
U F6RT34 F6RT34_MOUSE	27	322	298	313	319	20	48	1358	RNA processing ribosomal small subunit biogenesis translation	
U F6SIV1 F6SIV1_MOUSE	26	45	41	40	51	15	19	237	ATP synthesis coupled proton transport	
U F6XVMS F6XVMS_MOUSE	13	13	12	11	14	6	8	78	ATP synthesis coupled proton transport	
U F7D3R8 F7D3R8_MOUSE	16	18	20	15	22	8	12	111	ATP synthesis coupled proton transport	

tf D3YWB7 D3YWB7_MOUSE	4						48	74	126	actin filament-based process	
tf D3YX54 D3YX54_MOUSE	11	11	13	10	15	3	8	71	71	translation	Ribonucleoprotein Ribosomal protein
tf D3YXU1 D3YXU1_MOUSE	11	8	9	11	11	7	12	69	69	fatty acid beta-oxidation	
tf D3YV19 D3YV19_MOUSE	96	65	62	64	60	81	71	525	525	glucose metabolic process	Oxidoreductase
tf D3YZ35 D3YZ35_MOUSE	5	130	126	147	77	4	28	517	517	axon cargo transport intermediate filament bundle assembly microtubule cytoskeleton organization neurofilament cytoskeleton organization regulation of axon diameter	
tf D3YZ68 D3YZ68_MOUSE	2	3	5	3	3	2	1	56	56	GTP catabolic process	
tf D3YZV0 D3YZV0_MOUSE	20	20	24	18	22	20	25	162	162	ATP binding	
tf D3Z0L4 D3Z0L4_MOUSE	27	15	17	15	20	11	23	128	128	translation	
tf D3Z1N9 D3Z1N9_MOUSE	7	17	16	16	19	4	8	87	87	translation	

Tf A2AR91 A2AE91_MOUSE	7	7	8	9	7	5	10	53	nitrobenzene metabolic process xenobiotic catabolic process	
Tf A2AKU9 A2AKU9_MOUSE	33	36	32	18	24	15	30	189	ATP synthesis coupled proton transport	
Tf A2AQ9 A2AQ9_MOUSE	3					31	86	120	regulation of actin filament length sarcomere organization	
Tf A2AQB2 A2AQB2_MOUSE	4					18	42	64	regulation of actin filament length sarcomere organization	
Tf A2ARF5 A2ARF5_MOUSE	1	8	11	16	13		1	50	ATP binding kinase activity	Kinase Transferase
Tf A2ARF8 A2ARF8_MOUSE		67		67	75	93	1	303	microtubule cytoskeleton organization sensory perception of sound	
Tf A2BGC7 A2BGC7_MOUSE	19	8	9	9	15	3	8	71	In utero embryonic development negative regulation of striated muscle cell differentiation negative regulation of transcription from RNA polymerase II promoter	
Tf A6Z44 A6Z44_MOUSE	23	2	1	15	3	3	2	55	glycolysis response to estrogen stimulus response to heat response to hypoxia response to lipopolysaccharide response to nicotine	Lyase
Tf ABDUK4 ABDUK4_MOUSE	24	41	58	61	36	35	64	322	heme binding iron ion binding oxygen binding oxygen transporter activity	
Tf B0QZL1 B0QZL1_MOUSE	11	23	22	24	23	22	25	156	glycolysis	Lyase

tf AZAE91 AZAE91_MOUSE	7	7	8	9	7	5	10	53	nitrobenzene metabolic process xenobiotic catabolic process	
tf AZAKU9 AZAKU9_MOUSE	33	36	32	18	24	15	30	188	ATP synthesis coupled proton transport	
tf AZAQ9 AZAQ9_MOUSE	3					31	86	120	regulation of actin filament length sarcomere organization	
tf AZAQ92 AZAQ92_MOUSE	4					18	42	64	regulation of actin filament length sarcomere organization	
tf AZARP5 AZARP5_MOUSE	1	8	11	16	13		1	50	ATP binding kinase activity	Kinase Transferase
tf AZARP8 AZARP8_MOUSE		67	67	75	93	1		303	microtubule cytoskeleton organization sensory perception of sound	
tf AZBGC7 AZBGC7_MOUSE	19	8	9	9	15	3	8	71	in utero embryonic development negative regulation of striated muscle cell differentiation negative regulation of transcription from RNA polymerase II promoter	
tf AGZ44 AGZ44_MOUSE	23	2	1	15	3	3	2	55	glycolysis response to estrogen stimulus response to heat response to hypoxia response to lipopolysaccharide response to nicotine	Lyase
tf ABDUK4 ABDUK4_MOUSE	24	41	58	61	36	35	64	322	heme binding iron ion binding oxygen binding oxygen transporter activity	
tf BQZL1 BQZL1_MOUSE	11	23	22	24	23	22	25	156	glycolysis	Lyase

10. Bibliography

1. Tanaka F et al. (2012) Current status of treatment of spinal and bulbar muscular atrophy. *Neural Plast* 2012.
2. Zitzmann M, Nieschlag E (2003) The CAG repeat polymorphism within the androgen receptor gene and maleness. *Int J Androl* 26:76–83.
3. Sułek A et al. (2005) CAG repeat polymorphism in the androgen receptor (AR) gene of SBMA patients and a control group. *J Appl Genet* 46:237–239.
4. La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH (1991) Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* 352:77–79.
5. Garden GA, La Spada AR (2012) Intercellular (Mis)communication in Neurodegenerative Disease. *Neuron* 73:886–901.
6. Pradat P-F (2014) [SBMA: A rare disease but a classic ALS mimic syndrome]. *Presse Med* 43:580–6.
7. Hama T et al. (2012) Discrimination of spinal and bulbar muscular atrophy from amyotrophic lateral sclerosis using sensory nerve action potentials. *Muscle and Nerve* 45:169–174.
8. Funakoshi H et al. (1995) Muscle-derived neurotrophin-4 as an activity-dependent trophic signal for adult motor neurons. *Science* 268:1495–1499.
9. Katsuno M et al. (2012) Pathogenesis and therapy of spinal and bulbar muscular atrophy (SBMA). *Prog Neurobiol* 99:246–256.
10. Sorarù G et al. (2008) Spinal and bulbar muscular atrophy: Skeletal muscle pathology in male patients and heterozygous females. *J Neurol Sci* 264:100–105.
11. Ross CA (2002) Polyglutamine pathogenesis: Emergence of unifying mechanisms for Huntington's disease and related disorders. *Neuron* 35:819–822.
12. Ellerby LM et al. (1999) Kennedy's disease: Caspase cleavage of the androgen receptor is a crucial event in cytotoxicity. *J Neurochem* 72:185–195.
13. Kobayashi Y et al. (1998) Caspase-3 cleaves the expanded androgen receptor protein of spinal and bulbar muscular atrophy in a polyglutamine repeat length-dependent manner. *Biochem Biophys Res Commun* 252:145–150.
14. Li M et al. (1998) Nonneural nuclear inclusions of androgen receptor protein in spinal and bulbar muscular atrophy. *Am J Pathol* 153:695–701.
15. Adachi H et al. (2005) Widespread nuclear and cytoplasmic accumulation of mutant androgen receptor in SBMA patients. *Brain* 128:659–70.
16. Litvinov I V, De Marzo AM, Isaacs JT (2003) Is the Achilles' heel for prostate cancer therapy a gain of function in androgen receptor signaling? *J Clin Endocrinol Metab* 88:2972–82.
17. Li M et al. (1998) Nuclear inclusions of the androgen receptor protein in spinal and bulbar muscular atrophy. *Ann Neurol* 44:249–54.
18. La Spada AR, Taylor JP (2010) Repeat expansion disease: progress and puzzles in disease pathogenesis. *Nat Rev Genet* 11:247–258.
19. Zoghbi HY, Orr HT (2000) G LUTAMINE R EPEATS AND. *Annu Rev Neurosci* 23:217–247.
20. Zoghbi HY (2006) in *Principles of Molecular Medicine*, pp 1114–1122.
21. Zoghbi HY, Orr HT (2000) Glutamine repeats and neurodegeneration. *Annu Rev Neurosci* 23:217–247.
22. Trifiro MA, Kazemi-Esfarjani P, Pinsky L (1994) X-linked muscular atrophy and the androgen receptor. *Trends Endocrinol Metab* 5:416–21.
23. Schmidt BJ, Greenberg CR, Allingham-Hawkins DJ, Spriggs EL (2002) *Expression of X-linked bulbospinal muscular atrophy (Kennedy disease) in two homozygous women.*
24. Schöls L, Bauer P, Schmidt T, Schulte T, Riess O (2004) Autosomal dominant cerebellar ataxias: Clinical features, genetics, and pathogenesis. *Lancet Neurol* 3:291–304.
25. Orr HT (2012) Polyglutamine neurodegeneration: Expanded glutamines enhance native functions. *Curr Opin Genet Dev* 22:251–255.

26. Brinkman RR, Mezei MM, Theilmann J, Almqvist E, Hayden MR (1997) The likelihood of being affected with Huntington disease by a particular age, for a specific CAG size. *Am J Hum Genet* 60:1202–1210.
27. Wexler NS et al. (2004) Venezuelan kindreds reveal that genetic and environmental factors modulate Huntington's disease age of onset. *Proc Natl Acad Sci U S A* 101:3498–3503.
28. Gusella JF, MacDonald ME (2000) Molecular genetics: unmasking polyglutamine triggers in neurodegenerative disease. *Nat Rev Neurosci* 1:109–115.
29. Taylor JP et al. (2003) Aggresomes protect cells by enhancing the degradation of toxic polyglutamine-containing protein. *Hum Mol Genet* 12:749–757.
30. Takahashi T, Katada S, Onodera O (2010) Polyglutamine diseases: Where does toxicity come from? What is toxicity? Where are we going? *J Mol Cell Biol* 2:180–191.
31. Slow EJ et al. (2005) Absence of behavioral abnormalities and neurodegeneration in vivo despite widespread neuronal huntingtin inclusions. *Proc Natl Acad Sci U S A* 102:11402–11407.
32. Hackam AS et al. (1998) The influence of huntingtin protein size on nuclear localization and cellular toxicity. *J Cell Biol* 141:1097–1105.
33. Katsuno M, Adachi H, Inukai A, Sobue G (2003) Transgenic mouse models of spinal and bulbar muscular atrophy (SBMA). *Cytogenet Genome Res* 100:243–251.
34. Martín-Aparicio E et al. (2001) Proteasomal-dependent aggregate reversal and absence of cell death in a conditional mouse model of Huntington's disease. *J Neurosci* 21:8772–8781.
35. Heiser V et al. (2000) Inhibition of huntingtin fibrillogenesis by specific antibodies and small molecules: implications for Huntington's disease therapy. *Proc Natl Acad Sci U S A* 97:6739–6744.
36. Ehrnhoefer DE et al. (2006) Green tea (-)-epigallocatechin-gallate modulates early events in huntingtin misfolding and reduces toxicity in Huntington's disease models. *Hum Mol Genet* 15:2743–2751.
37. Sánchez I, Mahlke C, Yuan J (2003) Pivotal role of oligomerization in expanded polyglutamine neurodegenerative disorders. *Nature* 421:373–379.
38. Hageman J et al. (2010) A DNAJB Chaperone Subfamily with HDAC-Dependent Activities Suppresses Toxic Protein Aggregation. *Mol Cell* 37:355–369.
39. Wang AM et al. (2013) Activation of Hsp70 reduces neurotoxicity by promoting polyglutamine protein degradation. *Nat Chem Biol* 9:112–8.
40. Zijlstra MP et al. (2010) Levels of DNAJB family members (HSP40) correlate with disease onset in patients with spinocerebellar ataxia type 3. *Eur J Neurosci* 32:760–770.
41. Saudou F, Finkbeiner S, Devys D, Greenberg ME (1998) Huntingtin acts in the nucleus to induce apoptosis but death does not correlate with the formation of intranuclear inclusions. *Cell* 95:55–56.
42. Gutekunst CA et al. (1999) Nuclear and neuropil aggregates in Huntington's disease: relationship to neuropathology. *J Neurosci* 19:2522–2534.
43. Kim S, Nollen EAA, Kitagawa K, Bindokas VP, Morimoto RI (2002) Polyglutamine protein aggregates are dynamic. *Nat Cell Biol* 4:826–831.
44. Bucciantini M et al. (2002) Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* 416:507–511.
45. Fiumara F, Fioriti L, Kandel ER, Hendrickson WA (2010) Essential role of coiled coils for aggregation and activity of Q/N-rich prions and PolyQ proteins. *Cell* 143:1121–1135.
46. Schaefer MH, Wanker EE, Andrade-navarro MA (2012) Evolution and function of CAG / polyglutamine repeats in protein – protein interaction networks. 4273–4287.
47. Spyros Petrakis, Martin H. Schaefer EEW and, Andrade-Navarro MA, Petrakis S, Schaefer MH, Wanker EE (2013) Aggregation of polyQ-extended proteins is promoted by interaction with their natural coiled-coil partners. *Bioessays* 35:503–507.
48. Wytenbach A et al. (2000) Effects of heat shock, heat shock protein 40 (HDJ-2), and proteasome inhibition on protein aggregation in cellular models of Huntington's disease. *Proc Natl Acad Sci U S A* 97:2898–2903.
49. Venkatraman P, Wetzel R, Tanaka M, Nukina N, Goldberg AL (2004) Eukaryotic proteasomes cannot digest polyglutamine sequences and release them during degradation of polyglutamine-containing proteins. *Mol Cell* 14:95–104.

50. Ciryam P, Tartaglia G, Morimoto RI, Dobson CM, Vendruscolo M (2013) Widespread Aggregation and Neurodegenerative Diseases Are Associated with Supersaturated Proteins. *Cell Rep* 5:781–790.
51. Lieberman AP, Harmison G, Strand AD, Olson JM, Fischbeck KH (2002) Altered transcriptional regulation in cells expressing the expanded polyglutamine androgen receptor. *Hum Mol Genet* 11:1967–1976.
52. Luthi-Carter R et al. (2002) Polyglutamine and transcription: gene expression changes shared by DRPLA and Huntington’s disease mouse models reveal context-independent effects. *Hum Mol Genet* 11:1927–1937.
53. Luthi-Carter R et al. (2002) Dysregulation of gene expression in the R6/2 model of polyglutamine disease: parallel changes in muscle and brain. *Hum Mol Genet* 11:1911–1926.
54. Park SH et al. (2013) PolyQ proteins interfere with nuclear degradation of cytosolic proteins by sequestering the Sis1p chaperone. *Cell* 154:134–145.
55. Scarafone N et al. (2012) Amyloid-like fibril formation by polyQ proteins: a critical balance between the polyQ length and the constraints imposed by the host protein. *PLoS One* 7:e31253.
56. Duennwald ML, Jagadish S, Giorgini F, Muchowski PJ, Lindquist S (2006) A network of protein interactions determines polyglutamine toxicity. *Proc Natl Acad Sci U S A* 103:11051–11056.
57. Gunawardena S et al. (2003) Disruption of axonal transport by loss of huntingtin or expression of pathogenic polyQ proteins in *Drosophila*. *Neuron* 40:25–40.
58. Fu L, Gao YS, Sztul E (2005) Transcriptional repression and cell death induced by nuclear aggregates of non-polyglutamine protein. *Neurobiol Dis* 20:656–665.
59. Von Mikecz A (2009) PolyQ fibrillation in the cell nucleus: who’s bad? *Trends Cell Biol* 19:685–691.
60. Chan WM et al. (2011) Expanded polyglutamine domain possesses nuclear export activity which modulates subcellular localization and toxicity of polyQ disease protein via exportin-1. *Hum Mol Genet* 20:1738–1750.
61. McCampbell A et al. (2000) CREB-binding protein sequestration by expanded polyglutamine. *Hum Mol Genet* 9:2197–2202.
62. Jung J, Bonini N (2007) CREB-binding protein modulates repeat instability in a *Drosophila* model for polyQ disease. *Science* 315:1857–1859.
63. Haibing J, Nucifora FC, Ross CA, DeFranco DB (2003) Cell death triggered by polyglutamine-expanded huntingtin in a neuronal cell line is associated with degradation of CREB-binding protein. *Hum Mol Genet* 12:1–12.
64. Nucifora FC et al. (2001) Interference by huntingtin and atrophin-1 with cbp-mediated transcription leading to cellular toxicity. *Science* 291:2423–2428.
65. Millecamps S, Julien J-P (2013) Axonal transport deficits and neurodegenerative diseases. *Nat Rev Neurosci* 14:161–76.
66. Siu LT, Gordon T (2003) Mechanisms controlling axonal sprouting at the neuromuscular junction. *J Neurocytol* 32:961–974.
67. Lee W-CM, Yoshihara M, Littleton JT (2004) Cytoplasmic aggregates trap polyglutamine-containing proteins and block axonal transport in a *Drosophila* model of Huntington’s disease. *Proc Natl Acad Sci U S A* 101:3224–3229.
68. Orr AL et al. (2008) N-terminal mutant huntingtin associates with mitochondria and impairs mitochondrial trafficking. *J Neurosci* 28:2783–2792.
69. DiFiglia M et al. (1997) Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. *Science* 277:1990–1993.
70. Seidel K et al. (2010) Axonal inclusions in spinocerebellar ataxia type 3. *Acta Neuropathol* 120:449–460.
71. Cha JHJ (2007) Transcriptional signatures in Huntington’s disease. *Prog Neurobiol* 83:228–248.
72. Perez MK et al. (1998) Recruitment and the role of nuclear localization in polyglutamine-mediated aggregation. *J Cell Biol* 143:1457–1470.
73. Komatsu M et al. (2006) Loss of autophagy in the central nervous system causes neurodegeneration in mice. *Nature* 441:880–884.

74. Breuer P, Haacke A, Evert BO, Wüllner U (2010) Nuclear aggregation of polyglutamine-expanded ataxin-3: fragments escape the cytoplasmic quality control. *J Biol Chem* 285:6532–6537.
75. Huynh DP, Del Bigio MR, Ho DH, Pulst SM (1999) Expression of ataxin-2 in brains from normal individuals and patients with Alzheimer's disease and spinocerebellar ataxia 2. *Ann Neurol* 45:232–241.
76. Trushina E et al. (2003) Microtubule destabilization and nuclear entry are sequential steps leading to toxicity in Huntington's disease. *Proc Natl Acad Sci U S A* 100:12171–12176.
77. Paulson H (2003) Polyglutamine neurodegeneration: minding your Ps and Qs. *Nat Med* 9:825–826.
78. Paulson HL, Fischbeck KH (1996) Trinucleotide repeats in neurogenetic disorders. *Annu Rev Neurosci* 19:79–107.
79. Liu B, Hong J-S (2003) Role of microglia in inflammation-mediated neurodegenerative diseases: mechanisms and strategies for therapeutic intervention. *J Pharmacol Exp Ther* 304:1–7.
80. Lobsiger CS, Cleveland DW (2007) Glial cells as intrinsic components of non-cell-autonomous neurodegenerative disease. *Nat Neurosci* 10:1355–1360.
81. Dupuis L, Echaniz-Laguna A (2010) Skeletal muscle in motor neuron diseases: therapeutic target and delivery route for potential treatments. *Curr Drug Targets* 11:1250–1261.
82. Sopher BL et al. (2004) Androgen receptor YAC transgenic mice recapitulate SBMA motor neuronopathy and implicate VEGF164 in the motor neuron degeneration. *Neuron* 41:687–699.
83. Johansen JA et al. (2009) Recovery of function in a myogenic mouse model of spinal bulbar muscular atrophy. *Neurobiol Dis* 34:113–120.
84. Monks DA et al. (2007) Overexpression of wild-type androgen receptor in muscle recapitulates polyglutamine disease. *Proc Natl Acad Sci U S A* 104:18259–18264.
85. Aguilaniu H, Gustafsson L, Rigoulet M, Nyström T (2003) Asymmetric inheritance of oxidatively damaged proteins during cytokinesis. *Science* 299:1751–1753.
86. Tokui K et al. (2009) 17-DMAG ameliorates polyglutamine-mediated motor neuron degeneration through well-preserved proteasome function in an SBMA model mouse. *Hum Mol Genet* 18:898–910.
87. Palazzolo I et al. (2009) Overexpression of IGF-1 in Muscle Attenuates Disease in a Mouse Model of Spinal and Bulbar Muscular Atrophy. *Neuron* 63:316–328.
88. Cortes CJ et al. (2014) Muscle expression of mutant androgen receptor accounts for systemic and motor neuron disease phenotypes in spinal and bulbar muscular atrophy. *Neuron* 82:295–307.
89. Lieberman AP et al. (2014) Peripheral androgen receptor gene suppression rescues disease in mouse models of spinal and bulbar muscular atrophy. *Cell Rep* 7:774–84.
90. Rinaldi C, Bott LC, Fischbeck KH (2014) Muscle matters in Kennedy's disease. *Neuron* 82:251–3.
91. Miyazaki Y et al. (2012) Viral delivery of miR-196a ameliorates the SBMA phenotype via the silencing of CELF2. *Nat Med* 18:1136–1141.
92. Katsuno M et al. (2002) Testosterone reduction prevents phenotypic expression in a transgenic mouse model of spinal and bulbar muscular atrophy. *Neuron* 35:843–854.
93. Yu Z et al. (2006) Androgen-dependent pathology demonstrates myopathic contribution to the Kennedy disease phenotype in a mouse knock-in model. *J Clin Invest* 116:2663–2672.
94. Katsuno M et al. (2003) Leuprorelin rescues polyglutamine-dependent phenotypes in a transgenic mouse model of spinal and bulbar muscular atrophy. *Nat Med* 9:768–73.
95. Banno H et al. (2009) Phase 2 trial of leuprorelin in patients with spinal and bulbar muscular atrophy. *Ann Neurol* 65:140–150.
96. Katsuno M et al. (2010) Efficacy and safety of leuprorelin in patients with spinal and bulbar muscular atrophy (JASMITT study): A multicentre, randomised, double-blind, placebo-controlled trial. *Lancet Neurol* 9:875–884.
97. La Spada AR et al. (1998) Androgen receptor YAC transgenic mice carrying CAG 45 alleles show trinucleotide repeat instability. *Hum Mol Genet* 7:959–967.
98. Monks DA et al. (2008) Androgen receptor and Kennedy disease/spinal bulbar muscular atrophy. *Horm Behav* 53:729–740.

99. Gelmann EP (2002) Molecular Biology of the Androgen Receptor. *J Clin Oncol* 20:3001–3015.
100. Quigley CA et al. (1995) Androgen receptor defects: Historical, clinical, and molecular perspectives. *Endocr Rev* 16:271–321.
101. Lubahn DB et al. (1988) Cloning of human androgen receptor complementary DNA and localization to the X chromosome. *Science* (80-) 240:327–330.
102. Chang CS, Kokontis J, Liao ST (1988) Molecular cloning of human and rat complementary DNA encoding androgen receptors. *Science* (80-) 240:324–326.
103. Lonergan PE, Tindall DJ (2011) Androgen receptor signaling in prostate cancer development and progression. *J Carcinog* 10:20.
104. Helsen C et al. (2011) Structural basis for nuclear hormone receptor DNA binding. *Mol Cell Endocrinol* 348:411–417.
105. Helsen C et al. (2012) Evidence for DNA-Binding Domain-Ligand-Binding Domain Communications in the Androgen Receptor. *Mol Cell Biol* 32:3033–3043.
106. Heery DM, Kalkhoven E, Hoare S, Parker MG (1997) A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature* 387:733–736.
107. Van Royen ME, van Cappellen W a, de Vos C, Houtsmuller AB, Trapman J (2012) Stepwise androgen receptor dimerization. *J Cell Sci* 125:1970–9.
108. Van De Wijngaert DJ et al. (2006) Novel FXXFF and FXXMF motifs in androgen receptor cofactors mediate high affinity and specific interactions with the ligand-binding domain. *J Biol Chem* 281:19407–19416.
109. Schaufele F et al. (2005) The structural basis of androgen receptor activation: intramolecular and intermolecular amino-carboxy interactions. *Proc Natl Acad Sci U S A* 102:9802–7.
110. Brodie J, McEwan IJ (2005) Intra-domain communication between the N-terminal and DNA-binding domains of the androgen receptor: Modulation of androgen response element DNA binding. *J Mol Endocrinol* 34:603–615.
111. Heinlein CA, Chang C (2002) Androgen receptor (AR) coregulators: An overview. *Endocr Rev* 23:175–200.
112. Lavery DN, McEwan IJ (2008) Functional characterization of the native NH₂-terminal transactivation domain of the human androgen receptor: Binding kinetics for interactions with TFIIF and SRC-1a. *Biochemistry* 47:3352–3359.
113. He B, Lee LW, Minges JT, Wilson EM (2002) Dependence of selective gene activation on the androgen receptor NH₂- and COOH-terminal interaction. *J Biol Chem* 277:25631–25639.
114. Langley E, Zhou ZX, Wilson EM (1995) Evidence for an anti-parallel orientation of the ligand-activated human androgen receptor dimer. *J Biol Chem* 270:29983–29990.
115. He B et al. (2004) Structural basis for androgen receptor interdomain and coactivator interactions suggests a transition in nuclear receptor activation function dominance. *Mol Cell* 16:425–438.
116. He B, Wilson EM (2003) Electrostatic modulation in steroid receptor recruitment of LXXLL and FXXLF motifs. *Mol Cell Biol* 23:2135–2150.
117. Shen HC, Coetzee GA (2005) The Androgen Receptor: Unlocking the Secrets of Its Unique Transactivation Domain. *Vitam Horm* 71:301–319.
118. Kumar R et al. (2011) Role of the androgen receptor CAG repeat polymorphism in prostate cancer, and spinal and bulbar muscular atrophy. *Life Sci* 88:565–571.
119. Sircar K et al. (2007) Androgen receptor CAG repeat length contraction in diseased and non-diseased prostatic tissues. *Prostate Cancer Prostatic Dis* 10:360–368.
120. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN (2005) Flexible nets: The roles of intrinsic disorder in protein interaction networks. *FEBS J* 272:5129–5148.
121. Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 6:197–208.
122. Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* 293:321–331.
123. Babu MM, van der Lee R, de Groot NS, Gsponer J (2011) Intrinsically disordered proteins: Regulation and disease. *Curr Opin Struct Biol* 21:432–440.
124. Fuxreiter M et al. (2008) Malleable machines take shape in eukaryotic transcriptional regulation. *Nat Chem Biol* 4:728–737.
125. Uversky VN (2002) Natively unfolded proteins: a point where biology waits for physics. *Protein Sci* 11:739–756.

126. Lacy ER et al. (2004) p27 binds cyclin-CDK complexes through a sequential mechanism involving binding-induced protein folding. *Nat Struct Mol Biol* 11:358–64.
127. Schreiber G, Fersht AR (1993) Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. *Biochemistry* 32:5145–5150.
128. Lu Q, Lu HP, Wang J (2007) Exploring the Mechanism of Flexible Biomolecular Recognition with Single Molecule Dynamics. *Phys Rev Lett* 98.
129. McEwan IJ (2012) Intrinsic disorder in the androgen receptor: identification, characterisation and drugability. *Mol Biosyst* 8:82.
130. Reid J, Kelly SM, Watt K, Price NC, McEwan IJ (2002) Conformational analysis of the androgen receptor amino-terminal domain involved in transactivation. Influence of structure-stabilizing solutes and protein-protein interactions. *J Biol Chem* 277:20079–20086.
131. Felli IC, Pierattelli R, Tompa P NMR to characterize flexible, disordered protein states: towards expanding the structure – function paradigm A CON 2D NMR experiment to highlight the importance of NMR to achieve atomic resolution information on intrinsically disordered proteins.
132. Bermel W et al. (2012) Speeding up sequence specific assignment of IDPs. *J Biomol NMR* 53:293–301.
133. Bermel W et al. (2013) High-dimensionality ¹³C direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins. *J Biomol NMR* 57:353–61.
134. Bermel W, Felli IC, Kümmerle R, Pierattelli R (2008) ¹³C direct-detection biomolecular NMR. *Concepts Magn Reson Part A Bridg Educ Res* 32:183–200.
135. Kosol S, Contreras-Martos S, Cedeño C, Tompa P (2013) Structural characterization of intrinsically disordered proteins by NMR spectroscopy. *Molecules* 18:10802–28.
136. Salmon L et al. (2010) NMR characterization of long-range order in intrinsically disordered proteins. *J Am Chem Soc* 132:8407–8418.
137. Wishart DS, Sykes BD, Richards FM (1991) Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J Mol Biol* 222:311–333.
138. Wishart DS (2011) Interpreting protein chemical shift data. *Prog Nucl Magn Reson Spectrosc* 58:62–87.
139. Yao J, Dyson HJ, Wright PE (1997) Chemical shift dispersion and secondary structure prediction in unfolded and partly folded proteins. *FEBS Lett* 419:285–289.
140. Jensen MR, Salmon L, Nodet G, Blackledge M (2010) Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts. *J Am Chem Soc* 132:1270–1272.
141. Felli IC, Pierattelli R (2014) Novel methods based on ¹³C detection to study intrinsically disordered proteins. *J Magn Reson* 241:115–125.
142. Harrison PM (2006) Exhaustive assignment of compositional bias reveals universally prevalent biased regions: analysis of functional associations in human and Drosophila. *BMC Bioinformatics* 7:441.
143. Burkhard P, Stetefeld J, Strelkov S V. (2001) Coiled coils: A highly versatile protein folding motif. *Trends Cell Biol* 11:82–88.
144. Halfmann R et al. (2011) Opposing Effects of Glutamine and Asparagine Govern Prion Formation by Intrinsically Disordered Proteins. *Mol Cell* 43:72–84.
145. McDonnell a V, Jiang T, Keating a E, Berger B (2006) Paircoil2: improved prediction of coiled coils from sequence. *Bioinformatics* 22:356–8.
146. Parry DAD, Fraser RDB, Squire JM (2008) Fifty years of coiled-coils and alpha-helical bundles: a close relationship between sequence and structure. *J Struct Biol* 163:258–269.
147. Anfinsen CB (1973) Principles that govern the folding of protein chains. *Science* 181:223–230.
148. Young JC, Agashe VR, Siegers K, Hartl FU (2004) Pathways of chaperone-mediated protein folding in the cytosol. *Nat Rev Mol Cell Biol* 5:781–791.
149. Wickner S, Maurizi MR, Gottesman S (1999) Posttranslational quality control: folding, refolding, and degrading proteins. *Science* 286:1888–1893.
150. Kampinga HH, Craig EA (2010) The HSP70 chaperone machinery: J proteins as drivers of functional specificity. *Nat Rev Mol Cell Biol* 11:579–592.
151. Hartl FU, Hayer-Hartl M (2009) Converging concepts of protein folding in vitro and in vivo. *Nat Struct Mol Biol* 16:574–581.

152. Vos MJ, Kanon B, Kampinga HH (2009) HSPB7 is a SC35 speckle resident small heat shock protein. *Biochim Biophys Acta - Mol Cell Res* 1793:1343–1353.
153. Vos MJ, Hageman J, Carra S, Kampinga HH (2008) Structural and functional diversities between members of the human HSPB, HSPH, HSPA, and DNAJ chaperone families. *Biochemistry* 47:7001–7011.
154. Morimoto RI (2008) Proteotoxic stress and inducible chaperone networks in neurodegenerative disease and aging. *Genes Dev* 22:1427–1438.
155. Hartl FU (1996) Molecular chaperones in cellular protein folding. *Nature* 381:571–579.
156. Laufen T et al. (1999) Mechanism of regulation of hsp70 chaperones by DnaJ cochaperones. *Proc Natl Acad Sci U S A* 96:5452–5457.
157. Mayer MP, Bukau B (2005) Hsp70 chaperones: cellular functions and molecular mechanism. *Cell Mol Life Sci* 62:670–684.
158. Summers DW, Douglas PM, Ren HY, Cyr DM (2009) The type I Hsp40 Ydj1 utilizes a farnesyl moiety and zinc finger-like region to suppress prion toxicity. *J Biol Chem* 284:3628–3639.
159. Fan CY, Ren HY, Lee P, Caplan AJ, Cyr DM (2005) The type I Hsp40 zinc finger-like region is required for Hsp70 to capture non-native polypeptides from Ydj. *J Biol Chem* 280:695–702.
160. Qiu XB, Shao YM, Miao S, Wang L (2006) The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. *Cell Mol Life Sci* 63:2560–2570.
161. Chapple JP, van der Spuy J, Poopalasundaram S, Cheetham ME (2004) Neuronal DnaJ proteins Hsj1a and Hsj1b: a role in linking the Hsp70 chaperone machine to the ubiquitin-proteasome system? *Biochem Soc Trans* 32:640–642.
162. Bailey CK, Andriola IFM, Kampinga HH, Merry DE (2002) Molecular chaperones enhance the degradation of expanded polyglutamine repeat androgen receptor in a cellular model of spinal and bulbar muscular atrophy. *Hum Mol Genet* 11:515–523.
163. Adachi H et al. (2003) Heat shock protein 70 chaperone overexpression ameliorates phenotypes of the spinal and bulbar muscular atrophy transgenic mouse model by reducing nuclear-localized mutant androgen receptor protein. *J Neurosci* 23:2203–11.
164. Muchowski PJ et al. (2000) Hsp70 and hsp40 chaperones can inhibit self-assembly of polyglutamine proteins into amyloid-like fibrils. *Proc Natl Acad Sci U S A* 97:7841–7846.
165. Rujano MA, Kampinga HH, Salomons FA (2007) Modulation of polyglutamine inclusion formation by the Hsp70 chaperone machine. *Exp Cell Res* 313:3568–3578.
166. Chan HY, Warrick JM, Gray-Board GL, Paulson HL, Bonini NM (2000) Mechanisms of chaperone suppression of polyglutamine disease: selectivity, synergy and modulation of protein solubility in *Drosophila*. *Hum Mol Genet* 9:2811–2820.
167. Cyr DM (2008) Swapping Nucleotides, Tuning Hsp70. *Cell* 133:945–947.
168. Takayama S, Reed JC (2001) Molecular chaperone targeting and regulation by BAG family proteins. *Nat Cell Biol* 3:E237–E241.
169. Kabbage M, Dickman MB (2008) The BAG proteins: A ubiquitous family of chaperone regulators. *Cell Mol Life Sci* 65:1390–1402.
170. Takayama S, Xie Z, Reed JC (1999) An evolutionarily conserved family of Hsp70/Hsc70 molecular chaperone regulators. *J Biol Chem* 274:781–786.
171. Demand J The Ubiquitin-related BAG-1 Provides a Link between the Molecular Chaperones Hsc70 / Hsp70 and the Proteasome *.
172. Carra S, Seguin SJ, Lambert H, Landry J (2008) HspB8 chaperone activity toward poly(Q)-containing proteins depends on its association with Bag3, a stimulator of macroautophagy. *J Biol Chem* 283:1437–1444.
173. Balch WE, Morimoto RI, Dillin A, Kelly JW (2008) Adapting proteostasis for disease intervention. *Science* 319:916–919.
174. Söti C, Csermely P (2007) Protein stress and stress proteins: implications in aging and disease. *J Biosci* 32:511–515.
175. Takeuchi H et al. (2002) Hsp70 and Hsp40 improve neurite outgrowth and suppress intracytoplasmic aggregate formation in cultured neuronal cells expressing mutant SOD1. *Brain Res* 949:11–22.
176. Kobayashi Y et al. (2000) Chaperones Hsp70 and Hsp40 suppress aggregate formation and apoptosis in cultured neuronal cells expressing truncated androgen receptor protein with expanded polyglutamine tract. *J Biol Chem* 275:8772–8778.

177. Adachi H et al. (2007) CHIP overexpression reduces mutant androgen receptor protein and ameliorates phenotypes of the spinal and bulbar muscular atrophy transgenic mouse model. *J Neurosci* 27:5115–5126.
178. Pratt WB, Toft DO (2003) Regulation of signaling protein function and trafficking by the hsp90/hsp70-based chaperone machinery. *Exp Biol Med (Maywood)* 228:111–133.
179. Patury S, Miyata Y, Gestwicki JE (2009) Pharmacological targeting of the Hsp70 chaperone. *Curr Top Med Chem* 9:1337–1351.
180. Glover JR, Lindquist S (1998) Hsp104, Hsp70, and Hsp40: A novel chaperone system that rescues previously aggregated proteins. *Cell* 94:73–82.
181. Williams AJ, Knutson TM, Colomer Gould VF, Paulson HL (2009) In vivo suppression of polyglutamine neurotoxicity by C-terminus of Hsp70-interacting protein (CHIP) supports an aggregation model of pathogenesis. *Neurobiol Dis* 33:342–353.
182. Neidhardt FC, Bloch PL, Smith DF (1974) Culture medium for enterobacteria. *J Bacteriol* 119:736–747.
183. Ladbury JE, Chowdhry BZ (1996) Sensing the heat: the application of isothermal titration calorimetry to thermodynamic studies of biomolecular interactions. *Chem Biol* 3:791–801.
184. Wiseman T, Williston S, Brandts JF, Lin LN (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem* 179:131–137.
185. Doucleff M, Hatcher-Skeers M, Crane NJ (2011) in *Pocket guide to Biomolecular NMR*, pp 79–81.
186. Bieri M et al. (2011) Macromolecular NMR spectroscopy for the non-spectroscopist: Beyond macromolecular solution structure determination. *FEBS J* 278:704–715.
187. Kwan AH, Mobli M, Gooley PR, King GF, MacKay JP (2011) Macromolecular NMR spectroscopy for the non-spectroscopist. *FEBS J* 278:687–703.
188. Cavanagh J, Fairbrother W, Palmer III A, Rance M, Skelton N (2007) *Protein NMR spectroscopy. Principles and Practice*.
189. Kazimierzczuk K, Zawadzka A, Koźmiński W (2009) Narrow peaks and high dimensionalities: Exploiting the advantages of random sampling. *J Magn Reson* 197:219–228.
190. Felli IC, Pierattelli R (2012) Recent progress in NMR spectroscopy: toward the study of intrinsically disordered proteins of increasing size and complexity. *IUBMB Life* 64:473–81.
191. Wishart DS, Sykes BD (1994) The ¹³C chemical-shift index: a simple method for the identification of protein secondary structure using ¹³C chemical-shift data. *J Biomol NMR* 4:171–180.
192. Wirmer J, Wolfgang P, Schwalbe H (2006) Motional properties of unfolded ubiquitin: A model for a random coil protein. *J Biomol NMR* 35:175–186.
193. Dalgarno DC, Levine BA, Williams RJP (1983) Structural information from NMR secondary chemical shifts of peptide ?? C-H protons in proteins. *Biosci Rep* 3:443–452.
194. Logan TM, Thériault Y, Fesik SW (1994) Structural characterization of the FK506 binding protein unfolded in urea and guanidine hydrochloride. *J Mol Biol* 236:637–648.
195. Bermel W et al. (2009) H-start for exclusively heteronuclear NMR spectroscopy: The case of intrinsically disordered proteins. *J Magn Reson* 198:275–281.
196. RINALDI PL (1983) HETERONUCLEAR 2D-NOE SPECTROSCOPY. *J Am Chem Soc* 105:5167–5168.
197. Vranken WF et al. (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins Struct Funct Genet* 59:687–696.
198. Katsuno M et al. (2006) Pathogenesis, animal models and therapeutics in Spinal and bulbar muscular atrophy (SBMA). *Exp Neurol* 200:8–18.
199. Trotter Y et al. (1995) Cellular localization of the Huntington's disease protein and discrimination of the normal and mutated form. *Nat Genet* 10:104–110.
200. Limbourg A et al. (2009) Evaluation of postnatal arteriogenesis and angiogenesis in a mouse model of hind-limb ischemia. *Nat Protoc* 4:1737–1746.
201. Pinotsi D et al. (2014) Direct observation of heterogeneous amyloid fibril growth kinetics via two-color super-resolution microscopy. *Nano Lett* 14:339–345.
202. Van de Linde S et al. (2011) Direct stochastic optical reconstruction microscopy with standard fluorescent probes. *Nat Protoc* 6:991–1009.
203. Van de Linde S, Sauer M (2014) How to switch a fluorophore: from undesired blinking to controlled photoswitching. *Chem Soc Rev* 43:1076–87.

204. Ong HG, Wang J Study of Carbon Nanotube Based Devices Using Scanning Probe Microscope.
205. Mirkin SM (2007) Expandable DNA repeats and human disease. *Nature* 447:932–40.
206. Wells RD, Dere R, Hebert ML, Napierala M, Son LS (2005) Advances in mechanisms of genetic instability related to hereditary neurological diseases. *Nucleic Acids Res* 33:3785–98.
207. Adachi H et al. (2005) Widespread nuclear and cytoplasmic accumulation of mutant androgen receptor in SBMA patients. *Brain* 128:659–70.
208. Li M et al. (1998) Nuclear inclusions of the androgen receptor protein in spinal and bulbar muscular atrophy. *Ann Neurol* 44:249–54.
209. Thomas PS et al. (2006) Loss of endogenous androgen receptor protein accelerates motor neuron degeneration and accentuates androgen insensitivity in a mouse model of X-linked spinal and bulbar muscular atrophy. *Hum Mol Genet* 15:2225–2238.
210. Rhodes LE et al. (2009) Clinical features of spinal and bulbar muscular atrophy. *Brain* 132:3242–3251.
211. Merry DE, Kobayashi Y, Bailey CK, Taye a a, Fischbeck KH (1998) Cleavage, aggregation and toxicity of the expanded androgen receptor in spinal and bulbar muscular atrophy. *Hum Mol Genet* 7:693–701.
212. Fischbeck KH, Lieberman a, Bailey CK, Abel a, Merry DE (1999) Androgen receptor mutation in Kennedy’s disease. *Philos Trans R Soc Lond B Biol Sci* 354:1075–8.
213. Cortes CJ et al. (2014) Muscle expression of mutant androgen receptor accounts for systemic and motor neuron disease phenotypes in spinal and bulbar muscular atrophy. *Neuron* 82:295–307.
214. Lieberman AP et al. (2014) Peripheral androgen receptor gene suppression rescues disease in mouse models of spinal and bulbar muscular atrophy. *Cell Rep* 7:774–84.
215. McEwan IJ, Lavery D, Fischer K, Watt K (2007) Natural disordered sequences in the amino terminal domain of nuclear receptors: lessons from the androgen and glucocorticoid receptors. *Nucl Recept Signal* 5:e001.
216. Morley JF, Brignull HR, Weyers JJ, Morimoto RI (2002) The threshold for polyglutamine-expansion protein aggregation and cellular toxicity is dynamic and influenced by aging in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* 99:10417–10422.
217. Buchanan G et al. (2004) Structural and functional consequences of glutamine tract variation in the androgen receptor. *Hum Mol Genet* 13:1677–1692.
218. Kumar R et al. (2011) Role of the androgen receptor CAG repeat polymorphism in prostate cancer, and spinal and bulbar muscular atrophy. *Life Sci* 88:565–571.
219. Ryan CP, Crespi BJ (2013) Androgen receptor polyglutamine repeat number: Models of selection and disease susceptibility. *Evol Appl* 6:180–196.
220. Nedelsky NB et al. (2010) Native functions of the androgen receptor are essential to pathogenesis in a *Drosophila* model of spinobulbar muscular atrophy. *Neuron* 67:936–52.
221. Renier KJ et al. (2014) Antiandrogen flutamide protects male mice from androgen-dependent toxicity in three models of spinal bulbar muscular atrophy. *Endocrinology* 155:2624–34.
222. Fernández-Rhodes LE et al. (2011) Efficacy and safety of dutasteride in patients with spinal and bulbar muscular atrophy: A randomised placebo-controlled trial. *Lancet Neurol* 10:140–147.
223. Kobayashi Y et al. (1998) Caspase-3 cleaves the expanded androgen receptor protein of spinal and bulbar muscular atrophy in a polyglutamine repeat length-dependent manner. *Biochem Biophys Res Commun* 252:145–150.
224. Ellerby LM et al. (1999) Kennedy’s disease: Caspase cleavage of the androgen receptor is a crucial event in cytotoxicity. *J Neurochem* 72:185–195.
225. Cohen SI a et al. (2013) Proliferation of amyloid- β 42 aggregates occurs through a secondary nucleation mechanism. *Proc Natl Acad Sci U S A* 110:9758–63.
226. Hellstrand E, Boland B, Walsh DM, Linse S (2010) Amyloid ??-protein aggregation produces highly reproducible kinetic data and occurs by a two-phase process. *ACS Chem Neurosci* 1:13–18.
227. Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J* 78:1606–1619.
228. McEwan IJ, Lavery D, Fischer K, Watt K (2007) Natural disordered sequences in the amino terminal domain of nuclear receptors: lessons from the androgen and glucocorticoid receptors. *Nucl Recept Signal* 5:e001.

229. Bermel W et al. (2012) Speeding up sequence specific assignment of IDPs. *J Biomol NMR* 53:293–301.
230. Felli IC, Piai A, Pierattelli R (2013) Recent advances in solution NMR studies: ¹³C direct detection for biomolecular NMR applications. *Annu Reports NMR Spectrosc* 80:359–418.
231. Piai A et al. (2014) “CON-CON” assignment strategy for highly flexible intrinsically disordered proteins. *J Biomol NMR* 60:209–18.
232. Tamiola K, Acar B, Mulder FAA (2010) Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J Am Chem Soc* 132:18000–18003.
233. He B, Kempainen JA, Wilson EM (2000) FXXLF and WXXLF sequences mediate the NH2-terminal interaction with the ligand binding domain of the androgen receptor. *J Biol Chem* 275:22986–94.
234. Marsh JA, Singh VK, Jia Z, Forman-Kay JD (2006) Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Sci* 15:2795–2804.
235. Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN (2010) PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta - Proteins Proteomics* 1804:996–1010.
236. Chai Y, Berke SS, Cohen RE, Paulson HL (2004) Poly-ubiquitin Binding by the Polyglutamine Disease Protein Ataxin-3 Links Its Normal Function to Protein Surveillance Pathways. *J Biol Chem* 279:3605–3611.
237. Cummings CJ et al. (2001) Over-expression of inducible HSP70 chaperone suppresses neuropathology and improves motor function in SCA1 mice. *Hum Mol Genet* 10:1511–1518.
238. Warrick JM et al. (1999) Suppression of polyglutamine-mediated neurodegeneration in Drosophila by the molecular chaperone HSP70. *Nat Genet* 23:425–428.
239. Stenoien DL et al. (1999) Polyglutamine-expanded androgen receptors form aggregates that sequester heat shock proteins, proteasome components and SRC-1, and are suppressed by the HDJ-2 chaperone. *Hum Mol Genet* 8:731–741.
240. Waelter S et al. (2001) Accumulation of Mutant Huntingtin Fragments in Aggresome-like Inclusion Bodies as a Result of Insufficient Protein Degradation. *Mol Biol Cell* 12:1393–1407.
241. Cummings CJ et al. (1998) Chaperone suppression of aggregation and altered subcellular proteasome localization imply protein misfolding in SCA1. *Nat Genet* 19:148–154.
242. Shorter J (2011) The mammalian disaggregase machinery: Hsp110 synergizes with Hsp70 and Hsp40 to catalyze protein disaggregation and reactivation in a cell-free system. *PLoS One* 6.
243. Young JE et al. (2007) Proteolytic cleavage of ataxin-7 by caspase-7 modulates cellular toxicity and transcriptional dysregulation. *J Biol Chem* 282:30150–30160.
244. Zoubeidi A et al. (2007) Cooperative interactions between androgen receptor (AR) and heat-shock protein 27 facilitate AR transcriptional activity. *Cancer Res* 67:10455–10465.
245. Ciocca DR, Fanelli MA, Cuello-Carrion FD, Castro GN (2010) Heat shock proteins in prostate cancer: from tumorigenesis to the clinic. *Int J Hyperthermia* 26:737–747.
246. Dekker N et al. (1993) Solution structure of the POU-specific DNA-binding domain of Oct-1. *Nature* 362:852–855.
247. Sette M et al. (1997) The structure of the translational initiation factor IF1 from E. coli contains an oligomer-binding motif. *EMBO J* 16:1436–1443.
248. Wüthrich K (2003) NMR studies of structure and function of biological macromolecules (Nobel lecture). *Angew Chem Int Ed Engl* 42:3340–3363.
249. Zuiderweg ERP (2002) Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry* 41:1–7.
250. 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* 375:377–382.
251. Hur E et al. (2004) Recognition and accommodation at the androgen receptor coactivator binding interface. *PLoS Biol* 2.
252. Van Durme J et al. (2009) Accurate prediction of DnaK-peptide binding via homology modelling and experimental data. *PLoS Comput Biol* 5.
253. Gao XC et al. (2012) The C-terminal helices of heat shock protein 70 are essential for J-domain binding and ATPase activation. *J Biol Chem* 287:6044–6052.

254. Fan C-Y, Lee S, Cyr DM (2003) Mechanisms for regulation of Hsp70 function by Hsp40. *Cell Stress Chaperones* 8:309–316.
255. Hartl FU, Hayer-Hartl M (2002) Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* 295:1852–1858.
256. Rüdiger S, Buchberger A, Bukau B (1997) Interaction of Hsp70 chaperones with substrates. *Nat Struct Biol* 4:342–349.
257. Liberek K, Marszalek J, Ang D, Georgopoulos C, Zylicz M (1991) Escherichia coli DnaJ and GrpE heat shock proteins jointly stimulate ATPase activity of DnaK. *Proc Natl Acad Sci U S A* 88:2874–2878.
258. Rauch JN, Gestwicki JE (2014) Binding of human nucleotide exchange factors to heat shock protein 70 (Hsp70) generates functionally distinct complexes in vitro. *J Biol Chem* 289:1402–1414.
259. Park S-H, Raines RT (2004) Fluorescence polarization assay to quantify protein-protein interactions. *Methods Mol Biol* 261:161–166.
260. Jameson DM, Seifried SE (1999) Quantification of protein-protein interactions using fluorescence polarization. *Methods* 19:222–233.
261. Rossi AM, Taylor CW (2011) Analysis of protein-ligand interactions by fluorescence polarization. *Nat Protoc* 6:365–387.
262. Zinzalla G, Thurston DE (2009) Targeting protein-protein interactions for therapeutic intervention: a challenge for the future. *Future Med Chem* 1:65–93.
263. Ivanov AA, Khuri FR, Fu H (2013) Targeting protein-protein interactions as an anticancer strategy. *Trends Pharmacol Sci* 34:393–400.
264. Hessenkemper W, Baniahmad a (2013) Targeting heat shock proteins in prostate cancer. *Curr Med Chem* 20:2731–40.
265. Karagöz GE et al. (2014) Hsp90-tau complex reveals molecular basis for specificity in chaperone action. *Cell* 156:963–974.
266. Ciocca DR, Fanelli MA, Cuello-Carrion FD, Castro GN (2010) Heat shock proteins in prostate cancer: from tumorigenesis to the clinic. *Int J Hyperthermia* 26:737–747.
267. Powers M V., Workman P (2007) Inhibitors of the heat shock response: Biology and pharmacology. *FEBS Lett* 581:3758–3769.
268. Jhaveri K, Taldone T, Modi S, Chiosis G (2012) Advances in the clinical development of heat shock protein 90 (Hsp90) inhibitors in cancers. *Biochim Biophys Acta* 1823:742–55.
269. Jego G, Hazoumé A, Seigneuric R, Garrido C (2013) Targeting heat shock proteins in cancer. *Cancer Lett* 332:275–285.
270. Rhodes LE et al. (2009) Clinical features of spinal and bulbar muscular atrophy. *J Clin Neurol* 25:285–287.
271. Everest SJ et al. (2006) No abnormal prion protein detected in the milk of cattle infected with the bovine spongiform encephalopathy agent. *J Gen Virol* 87:2433–2441.
272. Ruizeveld de Winter JA et al. (1991) Androgen receptor expression in human tissues: an immunohistochemical study. *J Histochem Cytochem* 39:927–936.
273. Herndon LA et al. (2002) Stochastic and genetic factors influence tissue-specific decline in ageing *C. elegans*. *Nature* 419:808–814.
274. Butterfield DA, Boyd-Kimball D, Castegna A (2003) Proteomics in Alzheimer's disease: Insights into potential mechanisms of neurodegeneration. *J Neurochem* 86:1313–1327.
275. Butterfield DA (2004) Proteomics: A new approach to investigate oxidative stress in Alzheimer's disease brain. *Brain Res* 1000:1–7.
276. Sowell RA, Owen JB, Allan Butterfield D (2009) Proteomics in animal models of Alzheimer's and Parkinson's diseases. *Ageing Res Rev* 8:1–17.
277. Lewis RA (2004) Medical phase contrast x-ray imaging: current status and future prospects. *Phys Med Biol* 49:3573–3583.
278. Zhou SA, Brahme A (2008) Development of phase-contrast X-ray imaging techniques and potential medical applications. *Phys Medica* 24:129–148.
279. Pfeiffer F, Weitkamp T, Bunk O, David C (2006) Phase retrieval and differential phase-contrast imaging with low-brilliance X-ray sources. *Nat Phys* 2:258–261.
280. Moosmann J et al. (2013) X-ray phase-contrast in vivo microtomography probes new aspects of *Xenopus* gastrulation. *Nature* 497:374–7.
281. Moosmann J et al. (2014) Time-lapse X-ray phase-contrast microtomography for in vivo imaging and analysis of morphogenesis. *Nat Protoc* 9:294–304.

282. Zehbe R et al. (2010) Going beyond histology. Synchrotron micro-computed tomography as a methodology for biological tissue characterization: from tissue morphology to individual cells. *J R Soc Interface* 7:49–59.
283. Lexell J, Taylor CC, Sjöström M (1988) What is the cause of the ageing atrophy? Total number, size and proportion of different fiber types studied in whole vastus lateralis muscle from 15- to 83-year-old men. *J Neurol Sci* 84:275–294.
284. Holloszy JO, Chen M, Cartee GD, Young JC (1991) Skeletal muscle atrophy in old rats: differential changes in the three fiber types. *MechAgeing Dev* 60:199–213.
285. Schiaffino S, Reggiani C (2011) Fiber types in mammalian skeletal muscles. *Physiol Rev* 91:1447–531.
286. Jackman RW, Kandarian SC (2004) The molecular basis of skeletal muscle atrophy. *Am J Physiol Cell Physiol* 287:C834–C843.
287. Montarras D et al. (2005) Direct isolation of satellite cells for skeletal muscle regeneration. *Science* 309:2064–2067.
288. Yin H, Price F, Rudnicki M a (2013) Satellite cells and the muscle stem cell niche. *Physiol Rev* 93:23–67.
289. Li Y, Lee Y i., Thompson WJ (2011) Changes in Aging Mouse Neuromuscular Junctions Are Explained by Degeneration and Regeneration of Muscle Fiber Segments at the Synapse. *J Neurosci* 31:14910–14919.
290. Lin S, Landmann L, Ruegg MA, Brenner HR (2008) The role of nerve- versus muscle-derived factors in mammalian neuromuscular junction formation. *J Neurosci* 28:3333–3340.
291. Witzemann V (2006) Development of the neuromuscular junction. *Cell Tissue Res* 326:263–271.
292. Murray LM, Talbot K, Gillingwater TH (2010) Review: Neuromuscular synaptic vulnerability in motor neurone disease: Amyotrophic lateral sclerosis and spinal muscular atrophy. *Neuropathol Appl Neurobiol* 36:133–156.
293. Jang YC, Van Remmen H (2011) Age-associated alterations of the neuromuscular junction. *Exp Gerontol* 46:193–198.
294. Wagner J et al. (2013) Anle138b: A novel oligomer modulator for disease-modifying therapy of neurodegenerative diseases such as prion and Parkinson’s disease. *Acta Neuropathol* 125:795–813.
295. Walsh DM, Klyubin I, Fadeeva J V, Rowan MJ, Selkoe DJ (2002) Amyloid-beta oligomers: their production, toxicity and therapeutic inhibition. *Biochem Soc Trans* 30:552–557.
296. Simoneau S et al. (2007) In vitro and in vivo neurotoxicity of prion protein oligomers. *PLoS Pathog* 3:1175–1186.
297. Winner B et al. (2011) In vivo demonstration that alpha-synuclein oligomers are toxic. *Proc Natl Acad Sci U S A* 108:4194–4199.
298. Dunker AK et al. (2001) Intrinsically disordered protein. *J Mol Graph Model* 19:26–59.
299. Albers M et al. (2005) Automated yeast two-hybrid screening for nuclear receptor-interacting proteins. *Mol Cell Proteomics* 4:205–213.
300. Jänne OA et al. (2000) Androgen-receptor-interacting nuclear proteins. *Biochem Soc Trans* 28:401–405.
301. Maere S, Heymans K, Kuiper M (2005) BiNGO: A Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. *Bioinformatics* 21:3448–3449.
302. Cline MS et al. (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc* 2:2366–2382.
303. Saito R et al. (2012) A travel guide to Cytoscape plugins. *Nat Methods* 9:1069–76.
304. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T (2011) Cytoscape 2.8: New features for data integration and network visualization. *Bioinformatics* 27:431–432.
305. Fuentealba RA et al. (2010) Interaction with polyglutamine aggregates reveals a Q/N-rich domain in TDP-43. *J Biol Chem* 285:26304–26314.
306. Orr HT et al. (1993) Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nat Genet* 4:221–226.
307. Brown KR, Jurisica I (2005) Online predicted human interaction database. *Bioinformatics* 21:2076–2082.
308. Serra HG et al. (2006) ROR α -Mediated Purkinje Cell Development Determines Disease Severity in Adult SCA1 Mice. *Cell* 127:697–708.

309. Harjes P, Wanker EE (2003) The hunt for huntingtin function: Interaction partners tell many different stories. *Trends Biochem Sci* 28:425–433.
310. Schmidt T et al. (2002) Protein surveillance machinery in brains with spinocerebellar ataxia type 3: Redistribution and differential recruitment of 26S proteasome subunits and chaperones to neuronal intranuclear inclusions. *Ann Neurol* 51:302–310.
311. Perutz MF, Johnson T, Suzuki M, Finch JT (1994) Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. *Proc Natl Acad Sci U S A* 91:5355–5358.
312. Perutz MF (1999) Glutamine repeats and neurodegenerative diseases: Molecular aspects. *Trends Biochem Sci* 24:58–63.
313. Culver BP et al. (2012) Proteomic analysis of wild-type and mutant huntingtin-associated proteins in mouse brains identifies unique interactions and involvement in protein synthesis. *J Biol Chem* 287:21599–21614.
314. Katsuno M, Adachi H, Tanaka F, Sobue G (2004) Spinal and bulbar muscular atrophy: Ligand-dependent pathogenesis and therapeutic perspectives. *J Mol Med* 82:298–307.
315. Masino L, Pastore A (2002) Glutamine repeats: structural hypotheses and neurodegeneration. *Biochem Soc Trans* 30:548–551.
316. Masino L et al. (2003) Domain architecture of the polyglutamine protein ataxin-3: a globular domain followed by a flexible tail. *FEBS Lett* 549:21–25.
317. Wetzel R (2012) Physical chemistry of polyglutamine: Intriguing tales of a monotonous sequence. *J Mol Biol* 421:466–490.
318. Menon RP et al. (2013) The Role of Interruptions in polyQ in the Pathology of SCA1. *PLoS Genet* 9.
319. Altschuler EL, Hud N V, Mazrimas JA, Rupp B (1997) Random coil conformation for extended polyglutamine stretches in aqueous soluble monomeric peptides. *J Pept Res* 50:73–75.
320. Lyu PC, Sherman JC, Chen A, Kallenbach NR (1991) Alpha-helix stabilization by natural and unnatural amino acids with alkyl side chains. *Proc Natl Acad Sci U S A* 88:5317–5320.
321. Jayaraman M et al. (2012) Slow amyloid nucleation via ??-helix-rich oligomeric intermediates in short polyglutamine-containing huntingtin fragments. *J Mol Biol* 415:881–899.
322. Duenwald ML, Jagadish S, Muchowski PJ, Lindquist S (2006) Flanking sequences profoundly alter polyglutamine toxicity in yeast. *Proc Natl Acad Sci U S A* 103:11045–11050.
323. Sivanandam VN et al. (2011) The aggregation-enhancing huntingtin N-terminus is helical in amyloid fibrils. *J Am Chem Soc* 133:4558–4566.
324. Castillo V, Chiti F, Ventura S (2013) The N-terminal Helix Controls the Transition between the Soluble and Amyloid States of an FF Domain. *PLoS One* 8.
325. Jayaraman M et al. (2012) Slow amyloid nucleation via ??-helix-rich oligomeric intermediates in short polyglutamine-containing huntingtin fragments. *J Mol Biol* 415:881–899.
326. Lakhani V V., Ding F, Dokholyan N V. (2010) Polyglutamine induced misfolding of Huntingtin Exon1 is modulated by the flanking sequences. *PLoS Comput Biol* 6.
327. Kar K, Jayaraman M, Sahoo B, Kodali R, Wetzel R (2011) Critical nucleus size for disease-related polyglutamine aggregation is repeat-length dependent. *Nat Struct Mol Biol* 18:328–336.
328. Jayaraman M et al. (2012) Kinetically competing huntingtin aggregation pathways control amyloid polymorphism and properties. *Biochemistry* 51:2706–2716.
329. Schaefer MH, Wanker EE, Andrade-navarro MA (2012) Evolution and function of CAG / polyglutamine repeats in protein – protein interaction networks. 4273–4287.
330. Spyros Petrakis, Martin H. Schaefer EEW and, Andrade-Navarro MA (2013) Aggregation of polyQ-extended proteins is promoted by interaction with their natural coiled-coil partners. *Bioessays* 35:503–507.
331. Witt SN (2010) Hsp70 molecular chaperones and Parkinson’s disease. *Biopolymers* 93:218–228.
332. Dunker a K, Gough J (2011) Sequences and topology: intrinsic disorder in the evolving universe of protein structure. *Curr Opin Struct Biol* 21:379–81.

333. Jochum T et al. (2012) Toxic and non-toxic aggregates from the SBMA and normal forms of androgen receptor have distinct oligomeric structures. *Biochim Biophys Acta - Mol Basis Dis* 1822:1070–1078.
334. Renier KJ et al. (2014) Antiandrogen flutamide protects male mice from androgen-dependent toxicity in three models of spinal bulbar muscular atrophy. *Endocrinology* 155:2624–2634.
335. Chen C, Fischbeck KH (2006) in *Genetic Instabilities and Neurological Diseases, Second Edition*, pp 211–220.
336. Grunseich C, Rinaldi C, Fischbeck K (2014) Spinal and bulbar muscular atrophy: Pathogenesis and clinical management. *Oral Dis* 20:6–9.
337. Hua Y, Vickers TA, Okunola HL, Bennett CF, Krainer AR (2008) Antisense Masking of an hnRNP A1/A2 Intronic Splicing Silencer Corrects SMN2 Splicing in Transgenic Mice. *Am J Hum Genet* 82:834–848.
338. Martinez TL et al. (2012) Survival Motor Neuron Protein in Motor Neurons Determines Synaptic Integrity in Spinal Muscular Atrophy. *J Neurosci* 32:8703–8715.
339. Rinaldi C et al. (2012) Insulinlike growth factor (IGF)-1 administration ameliorates disease manifestations in a mouse model of spinal and bulbar muscular atrophy. *Mol Med* 18:1261–8.
340. Nallamsetty S, Austin BP, Penrose KJ, Waugh DS (2005) Gateway vectors for the production of combinatorially-tagged His6-MBP fusion proteins in the cytoplasm and periplasm of Escherichia coli. *Protein Sci* 14:2964–2971.
341. Tropea JE, Cherry S, Nallamsetty S, Bignon C, Waugh DS (2007) A generic method for the production of recombinant proteins in Escherichia coli using a dual hexahistidine-maltose-binding protein affinity tag. *Methods Mol Biol* 363:1–19.