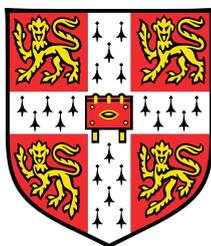


Proteomics studies of protein homeostasis and aggregation in ageing and neurodegeneration



Giulia Vecchi

Department of Chemistry

University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

Newnham College

February 2018

*Ai miei meravigliosi esseri ed alla mia dada e nonna,
che mi hanno insegnato a combattere, ad amare il sapere, a vivere ogni momento
e a non mollare mai*

To those marvellous beings that are my parents, and to my dada and granny,
who taught me to work hard, to love knowledge, to live to the fullest
and to never give up

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Giulia Vecchi

February 2018

Acknowledgements

These four years have passed very quickly and very slowly at the same time, with moments of high satisfaction alternated to bits of hard times as well. Overall, it has been one of the greatest experiences I lived so far, I believe mostly thanks to the people I have come to meet, interact and work with. Every moment has contributed to both my personal and professional growth, and there are many people I wish to send my deep thanks to for accompanying and supporting me in this important journey.

First, I would like to thank my supervisor, Prof Michele Vendruscolo, for giving me the opportunity to undertake this journey in the group and carry out this project. His expert guidance, encouragement and support were fundamental to my scientific and personal development over these years, so I am deeply thankful for the time he has dedicated to our discussions.

I am deeply grateful to Dr. Benedetta Mannini, who has been a wonderful colleague and a guide. She started the use of proteomics in the group and did all the experimental work for our projects. Her deep knowledge of biology has been fundamental for the interpretation of my analyses, our discussions are always very insightful and lively, and her comments very helpful. Working alongside her has been a great and pleasant learning experience.

I would like to thank as well Dr. Pietro Sormanni, for teaching me some crucial analysis skills during the first two years of my PhD studies, for his help, feedback and collaboration, and for taking the time to explain the details of the s2D predictor he developed and I used for my studies.

I would like to acknowledge Prof F. Ulrich Hartl, for the time and expertise he has devoted to our collaborations and from which I developed my interest for the raising field of proteomics.

Equally, I would like to thank Dr. Bertrand Fabre for sharing with me his expertise in mass spectrometry and proteomics. Discussing with him sped up incredibly my learning in the field, and was always very enjoyable.

The 142a office where I spent these last four years, the adjacent 145, and the basement have been vibrant places to work in and develop friendships. Thanks to all the people of the Vendruscolo group there, these are the places I felt I happily belonged to. Dr. Priyanka Joshi, Dr. Predrag Kukic, Prof. Stefano Gianni have been wonderful office mates, and I am very thankful for the nice time we spent together and for the great, lively discussions. I would like to thank also Andrea Possenti, who joined the group and the office recently and has been brainstorming with me on the content of our projects and the details of the analyses. Thanks to his expertise, similar background, and nice friendly attitude, our discussions have been really productive and inspiring.

Many thanks to all the Vendruscolo people I haven't thanked so far, for giving me insightful comments and suggestions inside and outside group meetings, and for making the lab a great place to work in.

On a more personal side, I wish to send my thanks to Maya Wright, Chistian Smyrilly, Jennifer Bellamy, Alicia Krozer, Marta Baldrighi and all the girls of the Women Blues Volleyball team who played with me in the last four years. You have been the greatest teammates I have ever had, and playing together while representing the Cambridge University team at the national finals and student cup finals, and arriving top 6 in the whole United Kingdom was one of the most fulfilling moments in my little sport career, and one of the most exciting in my life. Thanks to you, I can confidently affirm I truly lived the Cambridge experience, boosted my fighting spirit against any challenge, improved my team work. Once a Blue, always a Blue.

Thanks to all my friends in Cambridge, who I learnt to know in these years and have become among the dearest people to me. You gave me support in tough

times and joyful happy moments. Laura, who is now in Italy who has always been a support and a guide. Marta, who has been giving me power and keeping me company during the writing of this thesis. Sean, who took the trouble to go through this thesis and prepare me with a mock VIVA. Johnny, Priyanka, Benedetta, Swapan, Andrea and Karen. Thank you all, my 118 and weirdos, for making Cambridge a great place to live.

Pietro, AJ, Michele, Datta, Simone, Aprile, Gabi, Rosie, Sam, Mattia, Tessa. I am grateful to have met you all.

Finally, thanks to my friends at home, Giorgio, Susanna, Silvia, Alessandro, Carolina, Manuela, my friends for my former University. Mostly, thank you so much to my family and my parents, who has constantly supported and loved me throughout these years. I love you too.

Last, thanks to you again, Benedetta, because you are the greatest person I had the luck to meet, a dearest friend, a brilliant scientist, and a model of inspiration. So long, and thanks for all the fish(-pasta-food)!

Acknowledgements

Questi quattro anni sono sembrati allo stesso tempo un attimo ed una vita, con momenti di grandi soddisfazioni alternati a periodi molto difficili. Alla fin fine, principalmente grazie a tutte le persone che ho incontrato e con cui ho interagito e lavorato, questo dottorato si é rivelato una delle esperienze piú importanti della mia vita, ad oggi.

Ogni singolo istante ha contribuito alla mia crescita personale o professionale, e ci sono veramente molte persone che voglio ringraziare di cuore per avermi accompagnato e sostenuto in questo grande cammino.

Prima di tutti, vorrei ringraziare il mio supervisor, Prof. Michele Vendruscolo, per avermi dato l'opportunità di intraprendere questo viaggio e affrontare nel gruppo questo progetto di ricerca. La sua guida ed il suo costante supporto in questi anni sono state fondamentali per il mio sviluppo sia scientifico che personale, e sono molto grata per tutto il tempo che ha dedicato alle nostre discussioni.

I miei piú sentiti ringraziamenti vanno alla Dr. Benedetta Mannini, che é stata la mia guida oltre che una collega eccezionale. é stata lei ad aprire il settore della proteomica nel gruppo, e si é occupata di tutto il lavoro sperimentale nei progetti in cui ho lavorato in sua collaborazione. La sua infinita conoscenza della biologia é stata fondamentale per l'interpretazione delle mie analisi, e le nostre discussioni sono sempre produttive ed interessanti grazie ai suoi commenti estremamente utili e mirati. Lavorare al suo fianco é stata un'esperienza di apprendimento incredibilmente piacevole e importante per me.

Vorrei ringraziare anche il Dr. Pietro Sormanni, per il suo aiuto, feedback e collaborazione, specialmente per avermi insegnato delle tecniche essenziali di

analisi dati durante i primi due anni di dottorato ed aver speso tempo nel spiegarmi i dettagli del suo predittore s2D che ho applicato nei miei studi.

Vorrei esprimere il mio riconoscimento anche al Prof. F. Ulrich Hartl, per tutto il tempo e la conoscenza che ha dedicato alla nostra collaborazione e dalla quale ho realizzato il mio interesse per il campo emergente della proteomica.

Allo stesso modo, vorrei ringraziare il Dr. Bertrand Fabre, per aver condiviso con me la sua expertise in spettrometria di massa e proteomica. Conversare con lui é sempre piacevole, ed ha velocizzato incredibilmente la mia curva di apprendimento nel settore.

L'ufficio 142a, dove ho trascorso le giornate in questi ultimi quattro anni, l'adiacente ufficio 145 ed il basement del dipartimento sono luoghi che non dimenticheró mai, dove le menti lavorative entrano in risonanza e da cui nascono le amicizie. Grazie a tutte le persone del gruppo Vendruscolo che ho incontrato al loro interno, questi luoghi mi hanno fatto sentire felicemente a casa. Dr. Priyanka Joshi, Dr. Predrag Kukic, Prof. Stefano Gianni sono stati compagni d'ufficio fantastici, sono veramente grata per tutto il tempo che abbiamo passato insieme. Vorrei ringraziare anche Andrea Possenti, arrivato da poco in ufficio e nel gruppo, per il brainstorming praticamente giornaliero che facciamo sui progetti, i risultati e perfino i dettagli delle analisi. La sua expertise, lo stesso background scientifico, ma soprattutto il suo atteggiamento positivo e propositivo hanno reso le nostre discussioni incredibilmente produttive e stimolanti.

Mille grazie a tutti i Vendruscoli che non ho ancora ringraziato, per avermi sempre dato consigli e suggerimenti, e per aver reso il laboratorio un luogo fantastico in cui lavorare.

Passando al lato personale, ci terrei a ringraziare di cuore Maya Wright, Christiana Smyrilly, Jennifer Bellamy, Alicia Krozer, Marta Baldrighi e tutte le ragazze Women Blues Volleyball che hanno giocato con me in questi quattro anni: siete state le migliori compagne che abbia mai avuto, e giocare insieme rappresentando l'Universitá di Cambridge alle finali nazionali e alle finali delle coppe studenti, raggiungendo il traguardo di sesta squadra piú forte in tutto il Regno Unito, é stata

una delle tappe piú soddisfacenti ed appaganti di tutta la mia mini carriera sportiva, e sicuramente uno dei momenti piú emozionanti della mia vita. Grazie a voi, posso affermare con convinzione di aver veramente vissuto la "Cambridge experience", sviluppato uno spirito combattivo in grado di affrontare ogni difficoltà, e imparato cosa significhi veramente il "lavoro di squadra". "Once a Blue, always a Blue"; Leone Blu é per sempre.

Grazie grazie grazie a tutti gli amici della Cambrigia, che ho imparato pian piano a conoscere e che mi hanno sostenuto nei periodi piú bui e donato una quotidianità piena di spensieratezza, risate e allegria, divenendo in questi anni tra le persone a me piú care. Laura, che ora é tornata in Italia ma é sempre stata un sostegno ed una guida. Marta, che mi ha accompagnato durante praticamente tutta la scrittura di questa tesi, tenendomi compagnia e dandomi forza. Sean, che si é preso la briga di leggere questa tesi e simulare una prova di VIVA. E poi Johnny, Priyanka, Benedetta, Swapan, Andrea e ancora Karen: grazie, coinquiline del 118 e "weirdos", avete reso Cambridge un posto speciale dove vivere.

Pietro, AJ, Michele, Datta, Simone, Aprile, Gabi, Rosie, Sam, Mattia, Tessa. Sono grata di avervi incontrati tutti!

Infine, grazie ai miei amici di casa, ormai di lunga data: Giorgio, Susanna, Silvia, Alessandro, Carolina, Manuela e tutti gli amici ferraresi. Ma soprattutto, un eterno grazie alla mia famiglia ed ai miei genitori, che mi sostengono ed amano sin dal giorno che son nata. Vi amo tanto anche io.

E poi, grazie mille ancora a te, Benedetta, per essere la persona piú incredibile che io abbia avuto la fortuna di incontrare, una delle amiche piú care, una scienziata brillante ed una fonte di ispirazione continua. Per ogni giorno, ogni istante, ogni attimo che mi é stato dato, grazie mille per il pesce (la pasta e tutto il resto!)

Abstract

Upon ageing, a progressive disruption of protein homeostasis often leads to extensive protein aggregation and neurodegeneration. It is therefore important to study at the proteome level the origins and consequences of such disruption, which so far have remained elusive. Addressing this problem has recently become possible by major advances in mass spectrometry-based (MS) proteomics, which allows the identifications and quantification of thousands of proteins in a variety of biological samples.

In the first part of this thesis, I analyse proteome-wide MS data for the nematode worm *C. elegans* upon ageing, in wild type (WT), long-lived and short-lived mutant strains. By comparing the total abundance and the soluble abundance for nearly 4000 proteins, I provide extensive evidence that proteins are expressed in adult worms at levels close to their solubility limits. With the use of sequence-based prediction tools, I then identify specific physico-chemical properties associated with this age-related protein homeostasis impairment. The results that I obtained reveal that the total intracellular protein content remains constant, in spite of the fact that the proteome undergoes wide remodeling upon ageing, resulting into severe protein homeostasis disruption and widespread protein aggregation. These results suggest a protein-dependent decrease in solubility associated with the protein homeostasis failure.

In the second part of the thesis, I determine and classify potential interactions of misfolded protein oligomers with other proteins. This phenomenon is widely believed to give rise to cytotoxicity, although the mechanisms by which this happens are not fully understood. To address this question, I process and analyse MS data from structurally different oligomers (toxic type A and nontoxic type B) of the protein HypF-N, incubated in vitro with proteins extracted from murine cell cultures. I find that more than 2500 proteins are pulled down with the misfolded oligomers. These results indicate that the two types of oligomers interact with the same pool of proteins and differ only in the degree of binding. Functional annotation analysis on the groups reveals a preference of the oligomers to bind proteins in specific biological pathways and categories, including in particular mitochondrial membrane proteins, RNA-binding proteins and molecular chaperones.

Overall, in this study I complement the powerful and high-throughput experimental approach of MS proteomics with bioinformatics analyses and prediction algorithms to define the physical, chemical and biological features of protein homeostasis disruption upon ageing and the interactome of misfolded oligomers.

Table of contents

List of figures	xxi
List of tables	xxiii
1 Quantifying protein homeostasis in ageing and neurodegeneration	1
1.1 Ageing and neurodegeneration	1
1.1.1 A definition of ageing	1
1.1.2 Impact of neurodegenerative diseases	2
1.1.3 Clinical features of neurodegenerative diseases	4
1.2 Protein homeostasis and aggregation	7
1.2.1 The role of protein homeostasis in ageing and neurodegeneration	7
1.2.2 The interplay of protein homeostasis and aggregation	9
1.2.3 Protein aggregation in neurodegeneration	10
1.2.4 The generality of the aggregation process	13
1.2.5 Oligomers toxicity in the amyloid hypothesis	15
1.3 Mass spectrometry-based proteomics	17
1.3.1 Tools to measure the molecular changes in ageing and neurodegeneration	17
1.3.2 Mass Spectrometry-based proteomics for the direct measure of proteomes and protein interactions	20
1.4 Overview	23

2	Protein homeostasis imbalance and widespread aggregation in ageing	25
	<i>C. elegans</i>	25
2.1	Summary	26
2.2	The Camsol Intrinsic method	27
2.3	Publication: Widespread Proteome Remodeling and Aggregation in Aging <i>C. elegans</i>	28
2.4	Importance of the analysis	53
3	Life Over the Solubility Edge in ageing <i>C. elegans</i>	55
3.1	Introduction	56
3.2	Results	57
3.2.1	Life on and over the solubility edge in adult <i>C. elegans</i> . . .	57
3.2.2	Life over the solubility edge: the relationship between ag- gregation, supersaturation and abundance of proteins	58
3.2.3	Aggregate proliferation in ageing <i>C. elegans</i>	62
3.2.4	Age-dependent aggregate proliferation in terms of supersat- uration, aggregation and abundance	64
3.2.5	Functional protein classes responsible for the age-dependent aggregate increase	67
3.3	Conclusions	70
3.4	Materials and Methods	71
3.4.1	Calculations of total, soluble and pellet abundances in the adult nematode	71
3.4.2	Calculation of total and aggregate load variation upon ageing	72
4	The interactome of misfolded protein oligomers	75
4.1	Introduction	76
4.2	Results	79
4.2.1	Reproducibility of MS data	79
4.2.2	The structural difference of the oligomers modulates the binding strength to proteins	81

4.2.3	The two types of oligomers interact with similar pools of proteins	83
4.2.4	The biological constituents of the misfolded oligomers interactome	87
4.2.5	Mitochondrial and ribosomal proteins are the preferential binders of misfolded oligomers	89
4.3	Conclusions	96
4.4	Materials and Methods	98
4.4.1	Sample preparation and MS run	98
4.4.2	MS data processing and analysis	100
4.4.3	Functional annotation enrichment analysis	103
5	Conclusions and Perspectives	113
	References	117

List of figures

1.1	Estimated population growth in 2000-2050	3
1.2	Incidence of Alzheimer's disease in the U.S. population	5
1.3	The protein states and protein homeostasis network (PN)	8
1.4	Cycle of cellular impairment in ageing and disease	10
1.5	Protein aggregates in brain patients affected with neurodegenerative disorders	11
1.6	Diagram of the energy landscape for the reactions of folding and aggregation	14
1.7	Oligomers are more toxic to cells than larger aggregates	16
1.8	Individual and functional-associated comparison of mRNA levels and protein abundances	19
1.9	Bottom-up proteomic workflow for shotgun proteomics	21
3.1	Validation of the Life on the Edge hypothesis in adult worms	59
3.2	Visualisation of the most supersaturated, aggregated and abundant proteins in the life over the edge solubility landscape	61
3.3	Evolution of total and pellet amounts in ageing worms	63
3.4	Contribution of supersaturation, aggregation and abundance to aggregate amounts in ageing worms	66
4.1	Schematic representation of the two structurally different HypF-N oligomer types	77
4.2	MS-based proteomic schematic workflow	78
4.3	Reproducibility of the MS data of the biological replicas	80

4.4	Difference in binding power of type A and type B oligomers . . .	82
4.5	Oligomers bind the same pool of microglia proteins	85
4.6	Functional enrichment of the proteins in the oligomers interactome	88
4.7	Difference of natural abundance of proteins in microglia is not sufficient to explain the abundance percentages in the interactome	90
4.8	Misfolded oligomers preferentially bind mitochondrial and ribosomal proteins	92
4.9	Scheme of the 2D annotation abundance enrichment analysis . .	110

List of tables

1.1	Some of the various human disorders associated with amyloidosis	12
3.1	List of proteins contributing the most to the aggregate proliferation	68

Chapter 1

Quantifying protein homeostasis in ageing and neurodegeneration

1.1 Ageing and neurodegeneration

1.1.1 A definition of ageing

Ageing has been described as a time-dependent functional decline associated with an increase in probability of death and a diminution in fertility affecting essentially all living organisms [1]. It is a very general phenomenon involving a variety of processes at different biological scales, from the macroscopic scale of the whole organism to the microscopic scales of cellular and molecular mechanisms [2]. At the organismal level, some of the most familiar age-related phenotypes exhibited in humans include the redistribution of body fat, greying and thinning of hair, loss of vigour, muscle and skin tone [3]. While describing the ageing features may appear straightforward, characterising the microscopic origins of the overall phenomenon and the connections among all the different components is a complex task [4, 5]. Over the last 30 years, ageing research has been greatly facilitated by the isolation of the first long-lived strains in *Caenorhabditis elegans* (*C. elegans*) [6] which demonstrated

that genetic pathways and biochemical processes conserved in evolution can control, to some extent, the rate of ageing [7–14].

It has so become increasingly recognised that such progressive loss of physiological integrity, impaired functioning and increased vulnerability to death characteristic of the ageing phenotype can be directly related to the presence at a microscopic level of a time-dependent accumulation of cellular and molecular damage. Ageing is thus considered the primary risk factor for the most significant human pathologies of our century, namely cancer, diabetes, cardiovascular disorders and neurodegenerative diseases [9, 16–32].

1.1.2 Impact of neurodegenerative diseases

Of all the pathological conditions mentioned above, neurodegenerative disorders, including Alzheimer's, Parkinson's and Huntington's diseases, typical of aged individuals, have grown to become the most debilitating, common and expensive medical conditions of our age [16, 23–25, 33–39]. By contrast, deaths from other major causes have decreased significantly due to medical advances and improved social and environmental conditions (see Fig. 1.2A). These improvements have led to a substantial increase in the average human lifespan [40]. As a result, the worldwide population aged 60+ is projected to more than double in size between 2015 and 2030 (see Fig. 1.1A), eventually reaching about 2.1 billion people (see Fig. 1.1), as presented in a 2015 United Nation report on world population ageing [15].

These ageing trends are not only restricted to the western wealthy areas. Increasing of the life expectancy is affecting the human population globally. Indeed, the biggest contribution to such an increase is estimated to come from the less developed regions [15]. Also, in the more developed areas, the fastest growth rate will come from the oldest people, those in age range over 80 years old, who will reach about 434 million in 2030, more than three-fold the number present in 2015 (Fig. 1.1).

Such an improvement in life expectancy, however, comes at a price. As neither treatments nor efficient diagnostic tools have been found yet for neurodegenerative

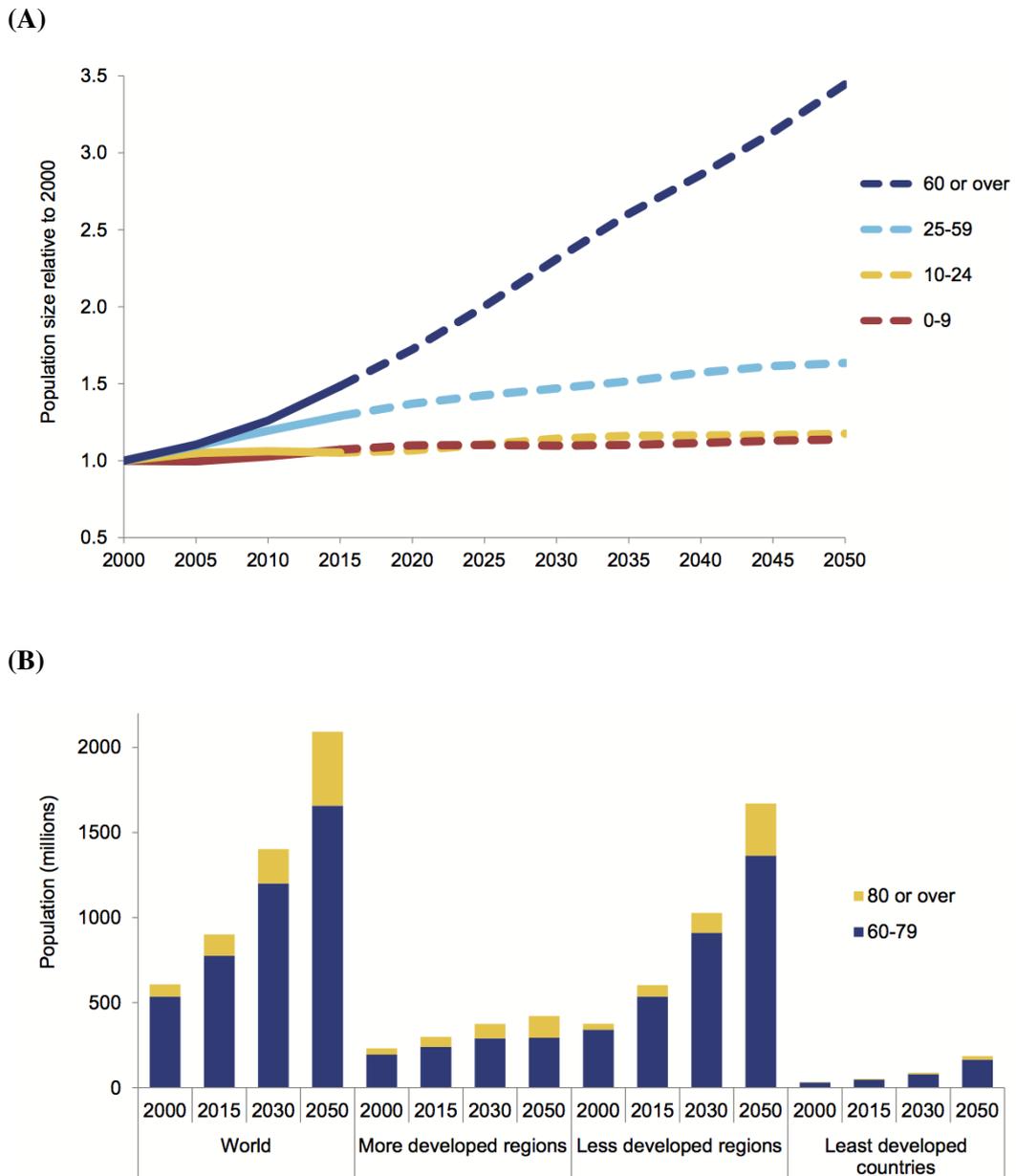


Fig. 1.1
Estimated population growth in 2000-2050

(A) Relative increase in world population by age group between 2000 and 2050. Solid lines represent calculated values, while dashed lines indicate projections. (B) Population size aged 60-79 years and 80+ years, separated by development group, calculated in 2000 and 2015, and predicted for 2030 and 2050.

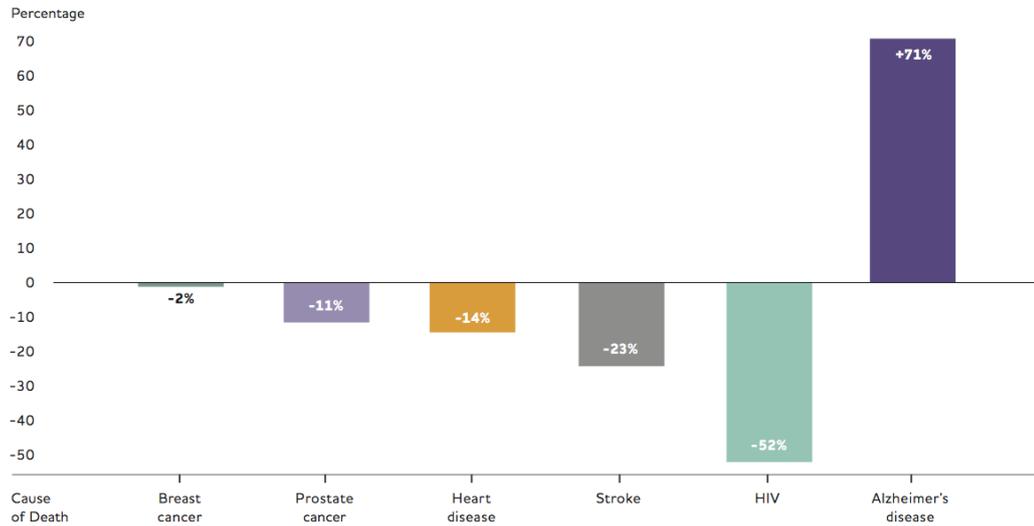
Figure reproduced from [15], see section 1.1.

disorders, these diseases have become the most dreaded pathologies of old age [16, 23–25, 33–39] not only in terms of mortality rate, but also in terms of social and economic burden. Indeed, Alzheimer's disease (AD) has been named "the twenty-first century plague" [35, 41] and is the most common cause of dementia, the condition of chronic cognitive decline that affect patient's day-to-day functions and spans multiple cognitive domains [36, 42–44]. In terms of mortality rates, American records [24] report that deaths from AD are the only kind of deaths that increased between 2000 and 2013, even massively (71%), compared to the current most common causes of death by disease (see Fig. 1.2A). In terms of economical and social burden, according to a 2011 economic report in the United Kingdom [43], the figures for dementia expenses to the UK economy in 2008, combining social care and health services, were about £10.5 billion, more than twice those for cancer (£4.5 billion), and over three fold those for stroke (£2.7 billion) and Coronary Heart Disease (£2.3 billion).

1.1.3 Clinical features of neurodegenerative diseases

Despite the great social and economic burden that dementia has been causing, this condition is still under-detected, under-diagnosed, under-treated and under-managed in primary care [39]. As already mentioned, the term dementia usually refers to a pathological state characterised by a progressive cognitive impairment, associated with a decline in memory, reasoning and language, resulting into severe issues in performing daily activities [25, 45]. Such impairment is the result of the damage and death of neurons in the regions of the brain responsible for cognitive function. This damage can be most often caused by four general categories of diseases: metabolic deficiencies (such as B12 thyroid hormone deficiencies), infections (such as syphilis and AIDS), structural lesions (such as neoplasm, tumours, and stroke), and neurodegeneration. However, since advanced diagnostic tests and early treatments currently exist for the first three processes but not for neurodegeneration, the vast majority

(A)



(B)

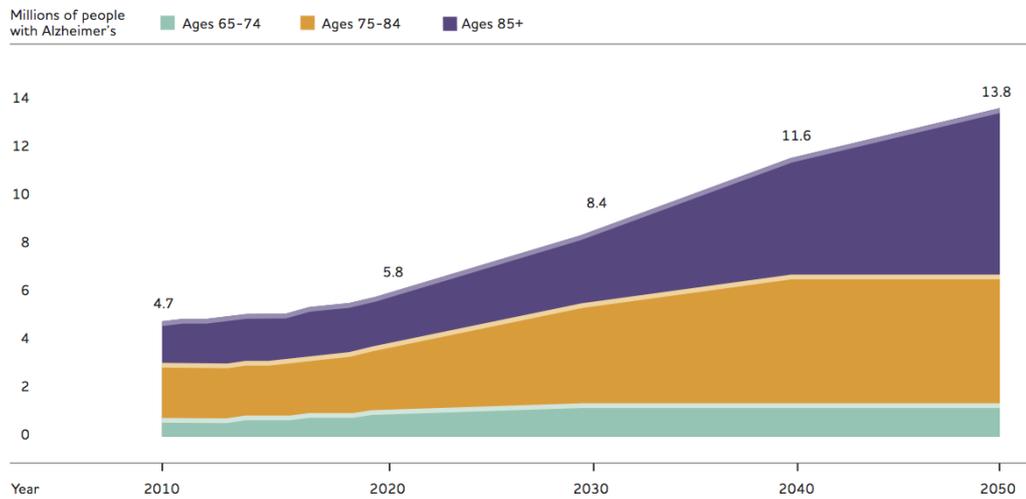


Fig. 1.2
Incidence of Alzheimer's disease in the U.S. population

(A) Percentage changes between 2000 and 2013 in the current most widespread pathological conditions. (B) Projected number of people in the US older than 65 suffering from Alzheimer's disease, from 2010 to 2050, grouped by age ranges.

Figure reproduced from [25], see section 1.1.

of patients suffering from dementia are those ending up having neurodegenerative diseases [36].

Diseases belonging to this class are varied in symptoms and characteristics [31, 46–55] as they target separate regions of the brain, especially in the early stages of the pathology. Mapping the anatomical sites of dysfunction is currently considered the most effective method for diagnosis. However, since there are no definitive markers available yet for the regional mapping of different neurodegenerative diseases, distinguishing among the different types of neurodegeneration remains a current problem at the clinical level [36]. In case of Alzheimer's disease, the most widespread among the neurodegenerative disorders [25, 35–38, 56], cognitive disability get combined with motility and swallowing problems, as neuronal damage and destruction progressively affect other parts of the brain responsible for basic bodily functions [57, 58]. As the transfer of information at the neuronal synapses starts to fail, the number of synapses decreases and neurons die. In a later stage, the increase of dead neurons leads to a dramatic shrinkage of the brain, and inflammation is also observed [25]. At the latest stage, the individual becomes bed-bound and eventually dies [25].

Despite the differences in the anatomical spreading patterns, brain regions involved and consequent symptoms, neurodegenerative diseases share a common attribute at the molecular level. While ageing is considered to be the greatest risk factor, a nearly universal hallmark of these disorders, and a common feature of other "non-neuronal diseases", like type II diabetes, is the formation and accumulation of protein aggregates, in the form of amyloid deposits [35, 59–63]. For this reason, such diseases have also been called protein conformational or misfolding diseases [59–62]. Description of this process and its relation with ageing, disease and cellular impairment will be discussed in section 1.2.3.

1.2 Protein homeostasis and aggregation

1.2.1 The role of protein homeostasis in ageing and neurodegeneration

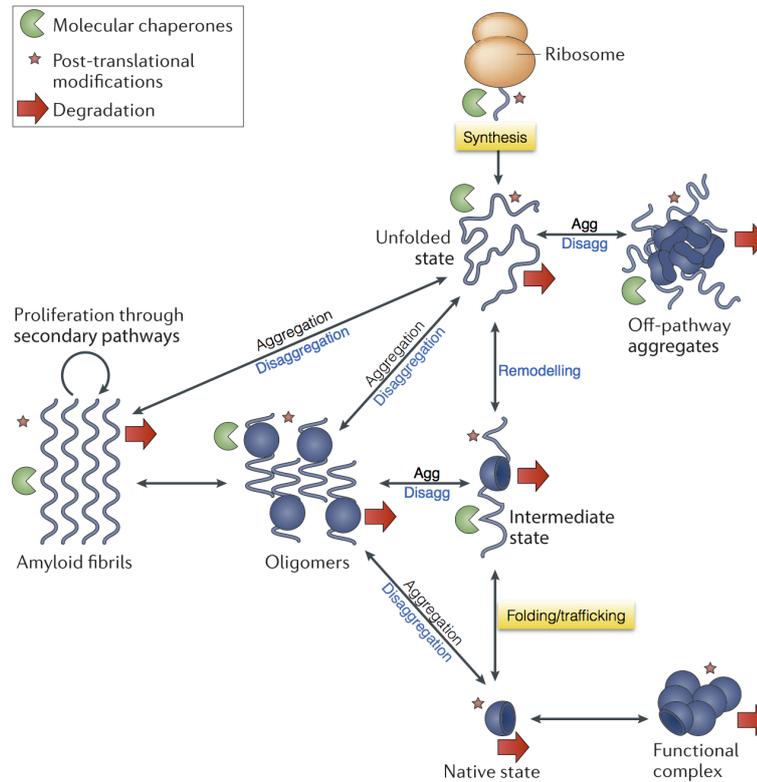
We have seen in section 1.1 how the ageing mechanism has recently gained the interest of the scientific community with the increase of the average human lifespan and concomitant diffusion of neurodegenerative disorders. Each level of biological organisation at which the ageing process can be defined is characterised by one or more physiological and cellular change, but the interplay between these processes and their causality relations are far from being fully understood [4]. The first attempt to define and categorise the main cellular hallmarks of ageing only dates to four years ago [9]. In that work, nine candidate hallmarks were proposed to contribute to the mammalian ageing process, and were functionally categorised as:

- primary cause of the cellular damage
- response to the damage
- subsequent reflection on the phenotype

Four physiological sources were suggested as primary cause of the cellular damage: genomic instability, telomere attrition, epigenetic alterations, and loss of protein homeostasis [64]. Of these four, the latter has emerged as an important factor also in neurodegenerative diseases [64–67]. Improving protein homeostasis through genetic manipulations has been shown to delay ageing in mammals [68, 69], while perturbation of protein homeostasis has been associated with ageing and fastening of age-associated pathologies [70–73].

Protein homeostasis refers to the state in which the proteome of a living organism is in functional balance, by the production and maintenance of correctly folded and soluble proteins [75, 76]. There are various biological mechanisms responsible for

(A)



(B)

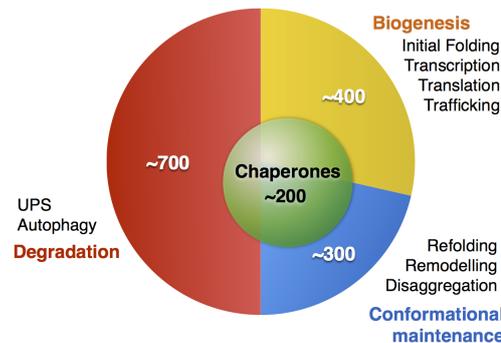


Fig. 1.3

The protein states and protein homeostasis network (PN)

(A) Diagram of the possible states in which a protein can exist coupled with the PN pathways which ensure the functional balance of the proteome through the three mechanisms of protein biogenesis (yellow components in the figure), maintenance (blue components in the figure) and degradation (red components in the figure) which control the levels of functional proteins and prevent or reduce the formation of toxic aggregates. (B) Approximate numbers of the PN components of the human proteome. 200 of the 1400 unities are molecular chaperones (shown in green).

Adapted from [67, 74] and [61].

protein homeostasis. Some have been conserved through evolution, while others have appeared more recently with the increasing complexity of cellular constituents, networks and organisms [74, 77, 78]. The main mechanisms responsible for the regulation of the folding state and the concentration distribution of the proteome can be grouped into three categories (see also Fig. 1.3):

- protein biogenesis, from transcription to translation
- protein maintenance, i.e. folding and refolding by molecular chaperones
- protein degradation, like autophagy and the proteasome system

These array of quality controls processes are necessary for cells to preserve the functionality and stability of their proteomes [70, 71, 79, 80].

1.2.2 The interplay of protein homeostasis and aggregation

Despite the great variety of such control mechanisms, as organisms age a progressive decline of protein homeostasis occurs as the cellular stress response and quality control become compromised [68, 73, 81].

In addition, disruption of protein homeostasis may occur when proteins fail to stay soluble and aggregate into non-functional and/or toxic species [82, 83]. This could produce an alteration of the normal protein homeostatic mechanism. One example is the failure of the proteasome in degrading insoluble protein aggregates and its subsequent loss of function from the association with the aggregates [84–86]. Also molecular chaperones, which are responsible for the protein maintenance machinery, have been seen to bound with aggregates in dynamic and complex ways [87–89]. Inhibiting these homeostatic mechanisms place a burden to the cell. This, in turn, can degenerate into a positive feedback loop in which the cell produce more proteins that aggregate, as they cannot be degraded nor refolded, and cause further inhibition of protein homeostasis (see Fig. 1.4, reproduced from [67]).

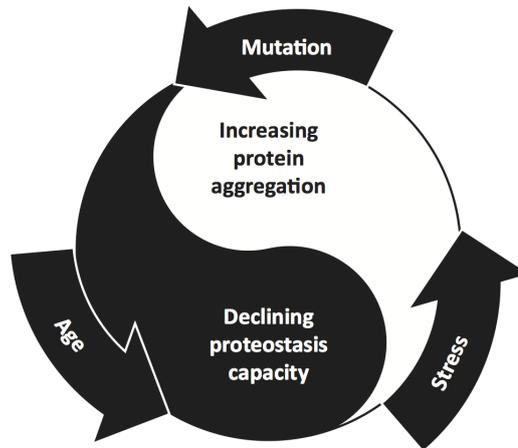


Fig. 1.4
Cycle of cellular impairment in ageing and disease

Aberrant misfolded species chronically produced with age, stress or disease place a burden to the protein homeostasis network (PN). Inhibition of this network in turns worsen the clearance of the non-functional/toxic aggregated species, which keep engaging the PN components impairing further the protein homeostasis and eventually leading to its collapse.

Figure reproduced from [67].

The features of protein homeostasis impairment and widespread aggregation occurring during the ageing process will be the subject of Chapter 2, a work performed from the collaboration with the group of Prof. F. Ulrich Hartl at the Max Planck Institute of Biochemistry in Martinsried, Germany [90].

1.2.3 Protein aggregation in neurodegeneration

Even if aggregation of non-disease related proteins has been recently observed to occur upon stress [91, 92] and during the ageing process [90, 93–95], the formation of protein aggregates and their implication in the failure of cellular machinery have first raised an interest in both the research and medical community because associated with a great variety of human diseases, from Type II diabetes and liver and skin amyloidosis to neurodegenerative disorders [96, 59–62, 97, 98]. These diseases are very diverse in symptoms and span over organ and tissue types [62], however they all share a common feature: the pathology involves the aggregation of one or

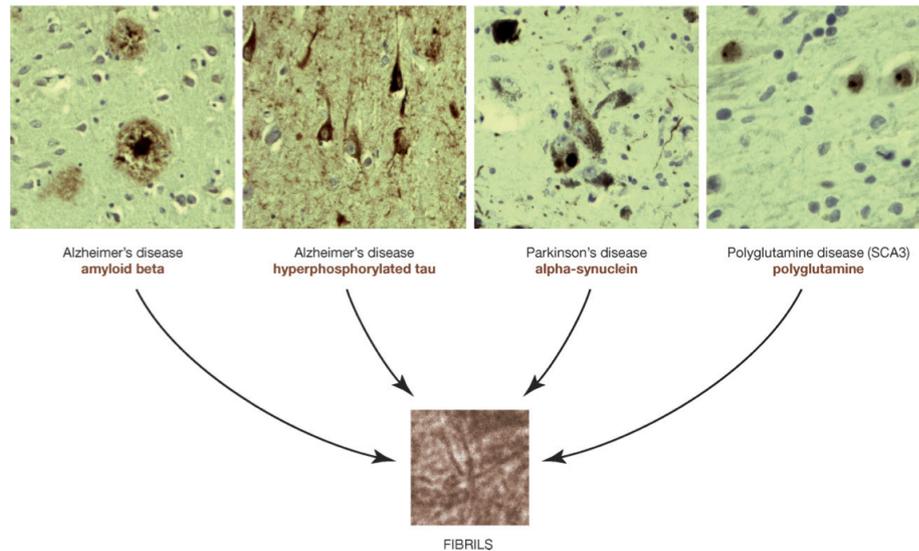


Fig. 1.5
Protein aggregates in brain patients affected with neurodegenerative disorders

Disease-specific proteins are shown in brown.
 Figure reproduced from [99].

more proteins, specific to the disease (see Fig. 1.5, reproduced from [99]). For this reason, diseases that seem so diverse have been grouped and classified as protein conformational or misfolding diseases. A list of the most widespread disorders among the various that fall under the category of misfolding diseases, together with the information on the characteristic aggregating protein, are found in Table 1.1.

The most notorious subset of disorders among the protein misfolding diseases is represented by the neurodegenerative pathologies like Alzheimer's disease (AD), Parkinson's disease (PD), and Huntington's disease (HD). In the case of AD, the accumulation of the protein *amyloid- β* outside neurons and of the protein *tau* intracellularly into amyloid aggregates, named respectively *amyloid- β* plaques and *tau* neurofibrillary tangles, represents the characteristic molecular fingerprint observed in the brains of patients affected by the disease. According to the "amyloid hypothesis" [59, 62], these aggregates are considered to contribute to the damage and destruction of neurons that result in memory loss and other symptoms of the pathology. While *amyloid- β* is believed to interfere in the inter-neurons communication at the

Table 1.1
Some of the various human disorders associated with amyloidosis

Adapted from [96].

Disease	Aggregating protein/peptide	Length
Neurodegenerative diseases		
Alzheimer's disease	<i>Amyloid-β</i> peptide	37-43
Creutzfeldt-Jakob diseases	Prion protein or its fragments	208
Gerstmann-Straussler-Scheinker disease	Prion protein or its fragments	208
Kuru	Prion protein or its fragments	208
Spongiform encephalopathies	Prion protein or its fragments	208
Huntington disease-like 1	Prion protein or its fragments	208
Parkinson's disease	<i>α-synuclein</i>	140
Dementia with Lewy bodies	<i>α-synuclein</i>	140
Multiple system atrophy	<i>α-synuclein</i>	140
Pick disease	Microtubule-associated protein tau	Variable
Progressive supranuclear palsy	Microtubule-associated protein tau	Variable
Corticobasal degeneration	Microtubule-associated protein tau	Variable
Argyrophilic grain disease	Microtubule-associated protein tau	Variable
Tangle predominant dementia	Microtubule-associated protein tau	Variable
Guam Parkinson dementia complex	Microtubule-associated protein tau	Variable
Frontotemporal lobar degeneration	Microtubule-associated protein tau	Variable
Chronic traumatic encephalopathy	Microtubule-associated protein tau	Variable
Hallervorden-Spatz disease	Microtubule-associated protein tau	Variable
Lipofuscinosis	Microtubule-associated protein tau	Variable
Amyotrophic lateral sclerosis	Superoxide dismutase 1	153
Huntington's disease	Huntingtin fragments	Variable
Familial amyloidotic polyneuropathy	Transthyretin mutants	127
Familial British dementia	ABri peptide	34
Familial Danish dementia	ADan peptide	34
Non-neuropathic systemic amyloidosis		
Light chain (AL) amyloidosis	Immunoglobulin light chains/fragments	~90
Heavy chain (AH) amyloidosis	Immunoglobulin heavy chains/fragments	~190
AA amyloidosis	Serum amyloid A protein fragments	45-104
Senile systemic amyloidosis	Wild-type transthyretin	127
Familial amyloidotic polyneuropathy	Wild-type transthyretin	127
Familial amyloid cardiomyopathy	Wild-type transthyretin	127
Dialysis-related amyloidosis	<i>β2-microglobulin</i>	99
Lysozyme amyloidosis	Lysozyme mutants	130
Apolipoprotein amyloidosis	Apo A-I, A-II, A-IV, C-II, C-III fragments	Variable
Familial finnish amyloidosis	gelsolin fragments	53 or 71
Non-neuropathic localized amyloidosis		
Type II diabetes	Islet amyloid polypeptide	37
Insulinoma	Islet amyloid polypeptide	37
Medullary carcinoma of the thyroid	Calcitonin	32
Atrial amyloidosis	Atrial natriuretic factor (ANF)	28
Pituitary prolactinoma	Prolactin fragments	34
Aortic medial amyloidosis	Medin	50
Gelatinous drop-like corneal dystrophy	Lactotransferrin	691
Injection-localized amyloidosis	Insulin	21 and 30

synapses, *tau* tangles seem to interfere with the intracellular transport of nutrients in the neurons, both processes contributing to cell death [25]. In the case of PD, the protein involved in the pathological aggregation is *α -synuclein*, a presynaptic protein 140 amino acids long, involved in vesicular transport [100–102]. Fibrillation of *α -synuclein* happens intracellularly in the dopaminergic neurons cell body, starting from brainstem structures (substantia nigra) and spreading to limbic and neocortical areas in the later stages of the disease [48, 46]. The formed aggregated structures are named Lewy bodies.

1.2.4 The generality of the aggregation process

Table 1.1 shows that each neurodegenerative disease involves the aggregation of one or more specific proteins. Despite the different nature of these proteins, the final structures in the aggregation process, the amyloid fibrils, have common characteristics that are independent of the constituting protein [103–106]. Amyloid structures appear as unbranched thread-like structures with diameters in the nanoscale and length ranging even up to micrometers, composed of multiple protofilaments twisted around each other [107]. They are highly ordered, closely packed, and possess a generic cross- β architecture [63, 108–112]. The cross- β structure allows the formation of a continuous array of hydrogen bonds along the length of the fibril, giving it high stability [107, 113, 114].

The observation that amyloid fibrils of different proteins possess this generic conserved architecture prompted the hypothesis that aggregation is a general mechanism that any polypeptide chain can be subject to. Evidence in support of this hypothesis has been reported both with *in vitro* and *in vivo* experiments. In the first case, it has been shown that almost every protein, even unrelated to diseases, can aggregate into amyloid fibrils under specific conditions [60, 115, 116]. In the latter case, work on *C. elegans* and yeast showed that many proteins unrelated to disease are found to form aggregates (widespread aggregation, see Chapter 2) with ageing or under stress conditions (e.g. starvation) [83, 93, 94, 117].

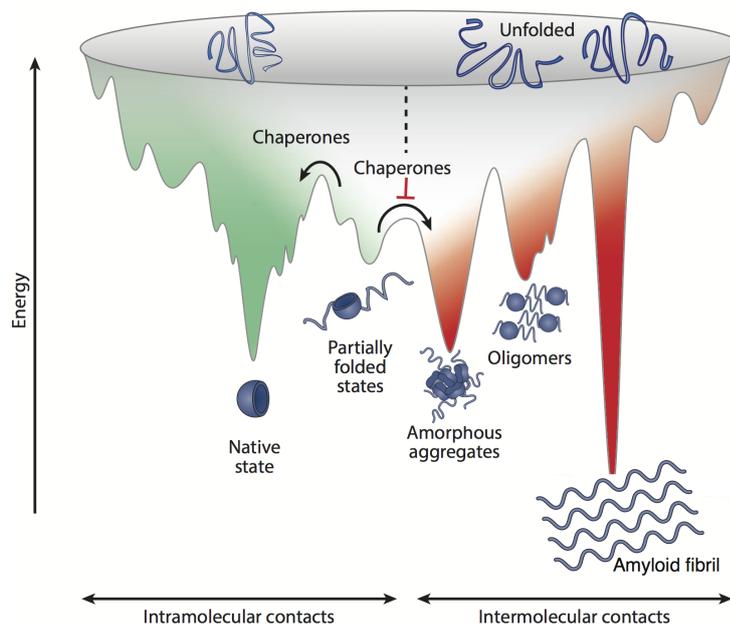


Fig. 1.6

Diagram of the energy landscape for the reactions of folding and aggregation

Competing reactions of protein folding and aggregation. A protein folds into its native functional state by sampling various intermediate conformations in a folding energy landscape. Energetically favourable intramolecular interactions (green) drive the protein progressing toward the native state, with the help of molecular chaperones. Molecular chaperones also try to avoid the formation of energetically favourable but nonnative conformations which happen when intermolecular interactions (red) occur and start the aggregation process. The amyloid fibrillar state resulting at the end of the aggregation process is the most thermodynamically stable. Adapted from [67, 74] and [61].

The hypothesis that every protein, under given conditions, can aggregate into an amyloid fibrillar structure implies that the amyloid structure form is a physical state that any protein in principle can assume. Furthermore, it implies that under those specific conditions the amyloid state is more thermodynamically stable than the functional native state of the protein itself [60, 74, 118, 119]. Strikingly, increasing evidence is showing that for many protein the amyloid fibrillar conformation might be the state of lowest free energy even under physiological conditions [106, 119–121].

The conditions under which a protein is pushed out of its soluble native-like state and aggregates depend on both the difference in energy between the native and amyloid state, and the height of the kinetic energy barrier that needs to be

overcome in order for this transition to occur [83, 106, 120, 122–128]. The amyloid state is characterised mostly by the intermolecular interactions, in contrast to the soluble native state in which intramolecular interactions dominate [61]. Since the conversion of proteins from the soluble to the aggregated state involve the formation of intermolecular contacts, a key role for the thermodynamic stability of the amyloid state is played by protein concentration. Indeed, as protein concentration increases, the ability to form intermolecular contacts becomes higher and favours the thermodynamic stability of the amyloid state, since the native state can be to a good approximation considered independent of the concentration [61, 120, 129]. Eventually, a critical concentration will be reached over which the free energy of the amyloid state will be lower than that of the native state.

Some proteins, however, have been reported to exist *in vivo* at physiological concentrations higher than the critical concentration [76, 120, 130, 131]. In this scenario, either spontaneous aggregation occurs; or high kinetic energy barriers prevent the transition from the native to the amyloid state, and the native state becomes metastable [119–121]. *In vivo*, the kinetic barrier is enhanced by the presence of molecular chaperones and the protein homeostasis network, which help proteins staying in their native state and prevent them from aggregate [77, 79, 120, 132, 133] (see Fig. 1.6). Whether the critical concentration *in vivo* is a general threshold for every protein under given conditions or is a protein-dependent value will be discussed in Chapter 3, with the results obtained from the measurements of proteome-wide protein abundances in *C.elegans*.

1.2.5 Oligomers toxicity in the amyloid hypothesis

The conversion of a protein from its soluble native form to the amyloid state is a complex self-assembly mechanism which involves the formation of multiple precursor soluble and insoluble species [98, 120, 123, 134–139]. These intermediate structures differ in size and characteristics, and range from early-stages small heterogeneous oligomeric assemblies to protofibrillar structures, smaller in length and width than the

final amyloids but organised in the same cross- β structure [105, 140, 141] (Fig.1.3A and Fig.1.6). Fig.1.3A and Fig.1.6 also show that amorphous aggregates can be formed, as an alternative pathway to the amyloid state formation [142, 143].

Despite the fact that the current diagnosis of neurodegenerative diseases is based on the spatial spreading pattern of amyloid fibrils, and the amyloid formation has been associated with the onset of the pathological event, little is known about the specific mechanisms of toxicity of the aggregates in the pathology [144–146]. While in systemic amyloidoses the mass of large fibrils in vital organs seems to represent the primary cause of the disease [62], an increasing number of studies now identifies the oligomeric pre-fibrillar species as the main pathogenic agents in neurodegenerative diseases [62, 98, 115, 146, 147].

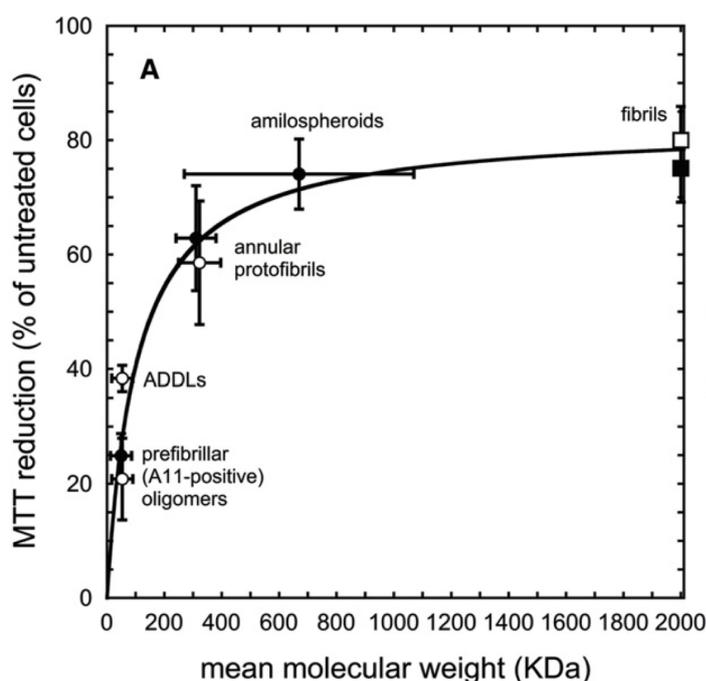


Fig. 1.7

Oligomers are more toxic to cells than larger aggregates

Toxicity versus size of $A\beta_{40}$ and $A\beta_{42}$ aggregates, measured by treating cultured cells with oligomers added to the extracellular medium and quantifying their MTT reduction relative to untreated cells. Aggregate toxicity is expressed as percentage of MTT reduction where the extremes, 0% and 100%, represent full cell death and full cell viability respectively. Small oligomeric species are found to compromise cell viability much stronger than fibrils and bigger aggregated species.

Figure reproduced from [148].

This hypothesis is the results of multiple observations: from the presence of many cases in which the amount of fibrillar aggregates does not correlate with the spreading of the disease [144, 145], to the evidence that oligomeric intermediates formed during the aggregation process are more toxic to cells than the mature amyloids (see Figure 1.7) [98, 115, 146–151]. Misfolded oligomers originated from proteins not associated with any pathology have also been demonstrated to bring damage to cells by making inappropriate interactions with lipid membranes and other functional cellular components [98, 150, 152]. Formation of larger structures like fibrils could be, in this context, a cellular response to the oligomeric damage. Indeed, we have detected the formation of bigger and more insoluble deposits have been detected in a long-lived strain of *C. elegans*. Results of this study, especially in terms of the physico-chemical principles of the insoluble aggregates, will be discussed in Chapter 2 and are found in [90].

In Chapter 4 we will present results on the interaction of misfolded oligomers that were found to mimic synaptic toxicity [152] with the functional cellular component that perform the vast majority of biological activities in the cell: the proteome. This results will provide insights in understanding how protein homeostasis can be potentially affected upon the presence of early-stages aggregated species in neurodegeneration.

1.3 Mass spectrometry-based proteomics

1.3.1 Tools to measure the molecular changes in ageing and neurodegeneration

So far, large-scale studies of the molecular factors associated with ageing and neurodegeneration have been mainly performed using genomics and transcriptomic approaches [4, 8, 153–157]. These studies tend to come from two different perspectives. On one side, they focus on finding features associated with longevity and

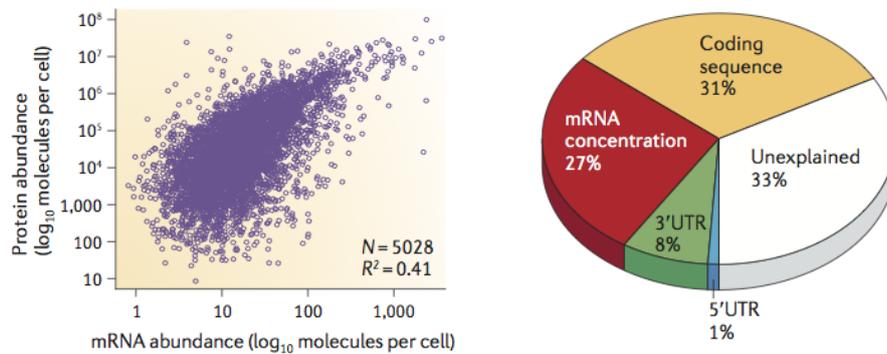
exceptional lifespan, predisposition and cause of disease, mostly in terms of genetic mutations. Genome-wide association studies (GWAS) have recently emerged as a powerful technique to address this task [154, 156, 158, 159]. On the other side, gene and mRNA expression analyses have become common practise to identify molecular changes that occur in healthy ageing and in neurodegeneration, either upstream as fingerprint of vulnerability or downstream as a response to the damage [155, 157, 160–162]. The systematic use of both techniques has also prompted the creation of numerous tools and databases aimed at integrating, collecting and extracting the biology and genetics of ageing and neurodegeneration from these large-scale information data [4, 157, 163, 164].

However, techniques that work at the gene and mRNA level provide only one type of information concerning what happens in a cell [165, 166]. In the cell, essentially all processes are catalysed and controlled by proteins, which account altogether for about 50% of the cell dry mass, with a concentration of more than 10^6 entities per cell volume [167].

Biophysics and biochemistry methods are conventionally used to isolate specific proteins and analyse their function and structure, but there are various ways in which proteins carry out their task and influence the behaviour of the cell in the cell: they can work at specific times, in precise locations, by themselves or in association with other proteins and molecules, organised in networks and structures. Collectively, proteins form a "proteome network" which determines the phenotype of the cell and is responsible for its functional state by dynamically adapting to external or internal perturbations, including genetic modifications [168].

Also, for any given protein a wide range of processes are required for its production and maintenance: from the transcription, processing and degradation of mRNAs, to the translation, localisation and degradation of the subsequent protein itself. The dynamic balance of these processes determines the cellular protein abundance at any given time, which in turns contributes to the balance of the proteome network and the maintenance of the protein homeostasis.

(A)



(B)

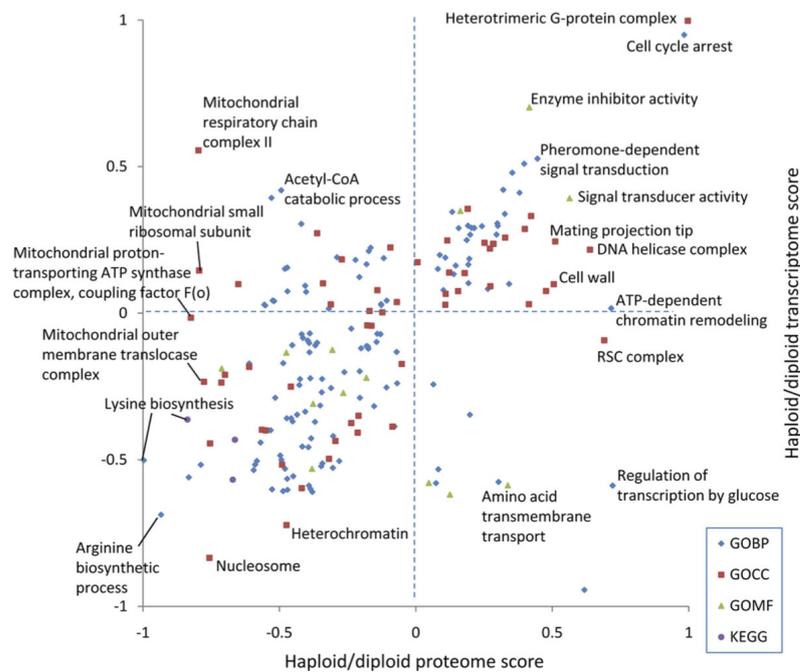


Fig. 1.8

Individual and functional-associated comparison of mRNA levels and protein abundances

(A) mRNA abundances do not correlate much with their respective proteins levels, as it is shown in the scatterplot on the left from data on mouse fibroblast cells where 5028 proteins and corresponding transcripts were quantified. On the right, a pie chart obtained from human DAOY medulloblastoma cell line shows that mRNA abundances can explain only for 30-40% of the variance in protein abundance.

Figure reproduced from [166]. (B) An example of scatterplot, from yeast cells proteomics and transcriptomics measurements, of the 2D-annotation enrichment [169] that quantitatively estimates significant upregulation or downregulation of entire biological processes and protein functional classes between the transcriptomic level and the proteomic level. Some processes belonging mostly to mitochondrial pathways are upregulated at the transcriptomic level compared to the proteomic level, while proteins belonging the amino acid transmembrane transport process and the glucose transport show an opposite behaviour. Figure reproduced from [169].

The importance of protein homeostasis and aggregation in ageing and neurodegeneration has thus prompted the need of large-scale studies of protein abundances and interactions. Since GWAS studies cannot monitor expression levels, the most common procedure so far consisted in using mRNA concentrations as proxies of activity and concentrations of the corresponding proteins. This choice follows the assumption that transcript levels are the main determinants of protein abundances. However, recent evidence obtained from actual measurements of protein abundances (see section 1.3.2) shows that for almost every organism tested, mRNA levels do not mirror cellular protein levels, but can only account for one to two third of their variations in the cell [166, 170, 171] (see Fig. 1.8A). This suggests that other processes beside and downstream mRNA translation, like post-transcriptional and translational regulation and protein degradation can give a dominant contribution to the regulation of protein expression levels. Moreover, the poor correlation between mRNA and protein abundances is not only restricted to single expression levels, but sometimes even involves the differential expression of entire biological pathways or functional components [169] (Figure 1.8B). For this reason, an accurate quantification of direct protein levels becomes fundamental for the study of protein homeostasis and protein interactions.

1.3.2 Mass Spectrometry-based proteomics for the direct measure of proteomes and protein interactions

In the last decades, major technological improvements in mass-spectrometry (MS)-based methods boosted the field of proteomics, the scientific area that aim to quantitatively study the proteome and the mechanisms that it mediates [172–192]. In terms of instrumentation, the development of the Orbitrap detector in 2005 [193] represented a milestone for the diffusion and affordability of rapid, high-sensitivity mass spectrometers, and permitted the nearly complete analysis of proteomes, an approach now addressed as "shotgun proteomics" [166, 181, 183, 188, 189, 194]. In such experiments, complete proteomes or proteins in a complex mixtures are enzymatically

digested and subsequently separated and analysed by nanoflow chromatography and high-resolution tandem mass spectrometry [172, 176, 179, 181, 183, 188, 195, 196]. This allows the identification of up of thousands proteins in a sample ("bottom-up proteomics").

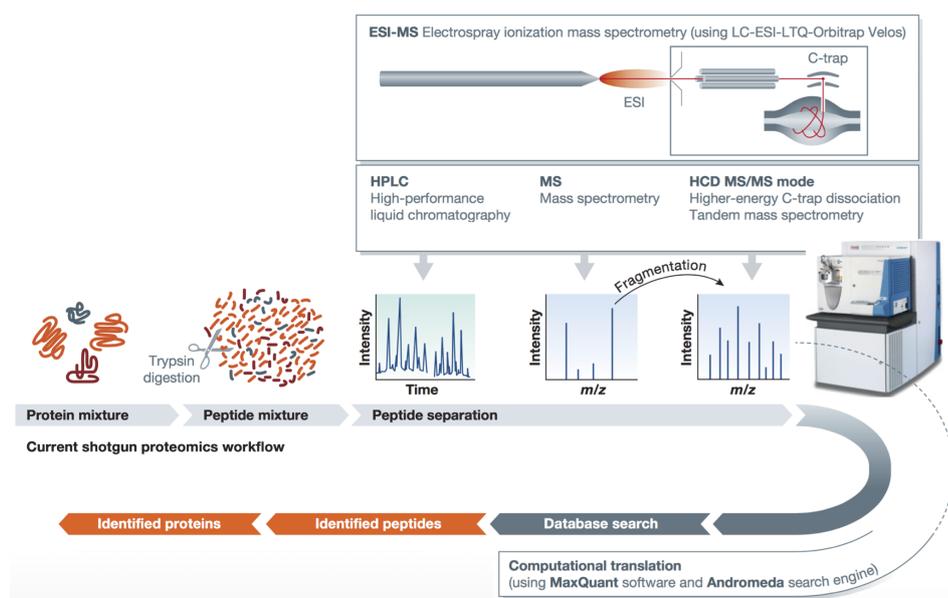


Fig. 1.9
Bottom-up proteomic workflow for shotgun proteomics

Proteins are first enzymatically digested, usually by trypsin protease, then separated by nanoflow chromatography and subsequently electrosprayed and analysed by high-resolution tandem mass spectrometry (MS), usually with an Orbitrap machine [193]. The first level of MS (MS1) acquires a full spectrum of the peptides that were eluting in the column in each recording time frame (MS run), while the second level (MS2) further fragments the most abundant MS1 peptides for identification of their amino acid sequence. Software packages can process the spectra and determine the protein content of the sample. Figure reproduced from [197].

A schematic representation of this MS-based proteomics workflow is shown in Fig. 1.9. The sample preparation is the first stage of the workflow, in which proteins are extracted from the sample and digested by a sequence-specific enzyme like trypsin. The peptides obtained from enzymatic digestion are then separated via liquid chromatography and electrosprayed before entering the vacuum of the mass spectrometer. As already mentioned, the most common mass spectrometer for bottom-up proteomics is the Orbitrap type mass analyser [193], which consists of

two MS levels: a first level (MS1) which allows the acquisition of the full spectrum of peptides eluting in the column with their intensities, followed by a second level (MS2) which collects as many fragmentation spectra as possible, within a cycle of 1-3 seconds, in order to identify the amino acid sequence of the most abundant peptides detected in the MS1 [168, 180, 193]. Resulting MS1 and MS2 spectra can be processed to obtain peptides identifications as well as protein identifications from the mapping of detected peptide sequences with a proteome sequence database.

This bottom-up proteomics approach has also been able to become quantitative [177, 198–201] with the development of data-processing software packages, which measure either spectral counts or ion intensities derived from each peptide, and reconstruct the signal of the parental protein. Among the softwares available that process the MS-MS data and measure ion intensities, which have been demonstrated to be more robust indicator of the protein levels in a sample than spectral counts [177, 202, 189, 203], the MaxQuant software has become highly recommended for proteome-wide protein quantification analyses, both in labelled or label-free conditions, thanks to his high peptide identification rates, mass accuracy and quantification technology [203–206].

In MaxQuant, peptide quantities are determined at the MS1 level by integrating the signal from peaks of the ions that elute from the chromatography column, with an alignment and normalisation step that allows comparison across MS runs. Also, a "match between runs" feature is present in the software to maximise the number of peptides identified and decrease the noise in their quantification, by transferring the information of peptide identity among runs in which the same aligned peak has been found (with same mass and elution time) but was not fragmented. The protein content of the sample is identified by mapping all detected and quantified peptides sequences to a database of protein sequences, which is the proteome (or protein mixture) of reference for the sample. The list of protein sequences in the reference is usually given as an input from the user, as it is experiment-dependent (e.g. in terms of organism or condition). In case of sequence degeneration, which occurs when a group of peptides is found to map multiple proteins (e.g. in case protein isoforms),

a single protein cannot be matched from the peptides and the ensemble of proteins containing the peptide sequences is reported (protein group). Absolute quantities for each protein (or protein group) can be estimated by summing up the intensities from the peak volumes of all peptides identifying the protein and normalising with respect to the number of theoretically expected peptides for that protein sequence upon *in silico* digestion with the enzyme used in the experiment: this method is called intensity-based absolute quantification (iBAQ) [203].

Overall, these recent advances in both instrumentation and downstream data analysis now show the capability of the MS-based proteomics technologies to both identify and accurately quantify proteins in large mixture, in either a labelled or label-free environment [207, 208], which represent a key step for the large-scale study of the proteome and of its interactions. In particular, it represents an invaluable resource for the investigation of protein homeostasis and widespread aggregation under various conditions, like ageing or neurodegeneration.

1.4 Overview

We described in section 1.2 the central role of protein homeostasis in ageing and neurodegeneration, and the hypothesis of misfolded oligomers toxicity in the neurodegenerative process. In this work, we studied these phenomena using MS-based proteomics coupled with sequence-based prediction algorithms and bioinformatic functional annotation analyses.

In Chapter 2, we will show how this approach can be applied to the identification of the features and physico-chemical principles responsible for age-related protein homeostasis impairment and widespread aggregation in the nematode *C.elegans*.

In Chapter 3, we will provide the first proteome-level support to the observation that proteins are expressed in the cells at their solubility limits. This hypothesis was firstly postulated by Tartaglia and co-workers based on a strong anticorrelation between the aggregation rates measured *in vitro* of the limited group of proteins

available in the literature at that time and the corresponding human mRNA expression levels measured *in vivo* [130].

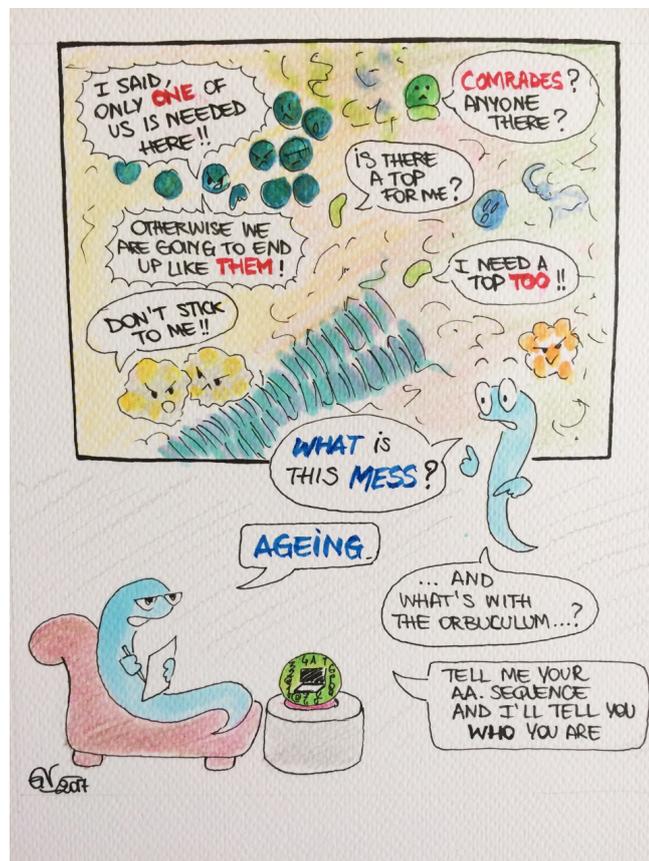
In Chapter 4, we perform MS proteomic analysis to identify and quantify the the interaction of proteins with misfolded oligomers, the potential most toxic species of the aggregation process. In this chapter, we couple the processing of MS data with bioinformatic functional annotation analysis to obtain biological insights into the interactions that oligomers can make, providing the functional oligomer interactome.

Both knowledge of the features of the proteome, its impairment with ageing, and its potential interaction with misfolded oligomers represent key questions for the understanding of neurodegeneration and may give rational insights for potential prevention or therapeutic approaches. Mass spectrometry and the recent advances in the technique are now allowing quantification of protein abundances *in vivo* on a proteome-wide scale, therefore this technique combined to our algorithms can be extremely useful in tracking upon aging the changes in the proteome system that result into protein homeostasis impairment.

Lastly, Chapter 5 shows the ensemble of results obtained from this work and the future perspectives that this study has opened.

Chapter 2

Protein homeostasis imbalance and widespread aggregation in ageing *C. elegans*



2.1 Summary

The content of this chapter is a published research article [90], which has been reprinted in section 2.3 with permission from Elsevier.

During my PhD studies, I contributed to the development of a computational method, called CamSol intrinsic, to determine the aggregation propensity of a protein/peptide, together with an array of other physico-chemical properties, from the information of its primary sequence only (see section 2.2). In this work, I use the CamSol intrinsic method and the s2D predictor [209] to understand physico-chemical principles underlying the age-related disruption of protein homeostasis and the widespread aggregation observed in nematode *C.elegans*.

The proteome in WT, long-lived and short-lived nematode strains was quantified upon ageing with MS in the group of Prof. F. Ulrich Hartl at the Max Planck Institute of Biochemistry, Germany. The main results of the analysis that I performed by complementing the biological information of the absolute abundances and relative changes in the both total protein concentration and the aggregated fraction with the results from the sequence-based physico-chemical and secondary structure calculations are shown in Figure 5 (D to G) and in Figure S5 (B and C) of the article (see section 2.3).

The analysis of physico-chemical properties from CamSol intrinsic and average coil propensity calculated from s2D was able to provide insightful information on the difference between the widespread aggregation observed in long-lived animals compared to short-lived animals. This, in turn, led to the first formulation of the hypothesis that the process of protein aggregation might be triggered by the cell as a mechanism for restoring protein homeostasis or delaying proteostasis impairment, a key novelty in the paper. Also, it revealed that the aggregation propensity of proteins is modulated by their abundance distribution in the cell, prompting the investigation of the proteome-wide evidence of the *life on the edge* hypothesis (see Chapter 3).

2.2 The Camsol Intrinsic method

Camsol Intrinsic is a computational method developed in this thesis in collaboration with Dr. Pietro Sormanni. Primarily, it is an extension of the Camsol method [210] tailored to the *in silico* sequence-based prediction of the aggregation propensities of proteins on a proteome-wide scale, but it also allows the calculation of an array of other key physico-chemical properties related.

There are two main reasons behind the choice of developing a sequence-based prediction method over a structure-based method for the analysis of proteome-wide data:

1. the amount of information available
2. the presence of intrinsically disordered proteins (IDPs) and proteins with intrinsically disordered regions (IDRs)

Regarding point 1, in fact, despite the great increase in the number of protein structures resolved in the recent years, the number of protein structures available is still an order of magnitude less than the number of protein sequences. To date, the number of all sequences present in the Uniprot database of proteins and peptides counts $\sim 10^9$ entries of which more than 550000 manually reviewed [211, 212], compared to the ~ 130000 structures stored in the Protein Data Bank (PDB) database [213, 214].

Also, it has become clear that not all proteins possess a stable structured fold but some are naturally structurally disordered (IDPs) or contain long stretches of disordered regions (IDRs) [215–218]. Such proteins play important functions in the cell, such as signalling and regulation [219–221]. Most importantly, more and more recent studies have highlighted that such proteins are extremely abundant in cells and especially in eukaryotic systems, where they make up $\sim 30\% - 45\%$ of the proteome [217, 220, 222–228]. Given that about a third of the proteome consists

of IDPs, the choice of choosing a sequence-based prediction method should allow consistent results to be obtained for high-throughput experimental data.

Given the amino-acid sequence, CamSol intrinsic evaluates the net charge, the fraction of charged residues, the average and total hydrophilicity and predicts an intrinsic aggregation propensity score, which represent the inherent predisposition of the protein/peptide to aggregate. In particular, CamSol intrinsic allows to evaluate the most/least vulnerable proteins of an ensemble (or a proteome) in terms of aggregation, by ranking them according to their CamSol intrinsic aggregation score (called zscore).

The method relies on the use the CamSol method [210] for the calculation of the residue-specific properties profiles from which the global properties are obtained with an optimisation strategy that allows comparison of the scores among different proteins of a proteome, unlike the CamSol method itself which is optimised for detecting differences in solubility from small variations in the sequence of a given protein [210], or the Zyggregator method [229] which was developed to specifically estimates amyloid aggregation rates of proteins. Details of the method are found in the supplementary information, section "In Silico Aggregation Propensity Calculation", in section 2.3.

2.3 Publication: Widespread Proteome Remodeling and Aggregation in Aging *C. elegans*

Reprinted from [90] with permission from Elsevier.

DOI: <http://dx.doi.org/10.1016/j.cell.2015.03.032>

Article

Cell

Widespread Proteome Remodeling and Aggregation in Aging *C. elegans*

Dirk M. Walther,^{1,5} Prasad Kasturi,^{2,5} Min Zheng,² Stefan Pinkert,^{2,7} Giulia Vecchi,³ Prajwal Ciryam,^{3,4}Richard I. Morimoto,⁴ Christopher M. Dobson,³ Michele Vendruscolo,³ Matthias Mann,^{1,6} and F. Ulrich Hartl^{2,6,*}¹Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany²Department of Cellular Biochemistry, Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany³Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK⁴Department of Molecular Biosciences, Rice Institute for Biomedical Research, Northwestern University, Evanston, IL 60208, USA⁵Co-first author⁶Co-senior author⁷Present address: Genomics and Proteomics Core Facility, German Cancer Research Center, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany*Correspondence: uhartl@biochem.mpg.de<http://dx.doi.org/10.1016/j.cell.2015.03.032>

SUMMARY

Aging has been associated with a progressive decline of proteostasis, but how this process affects proteome composition remains largely unexplored. Here, we profiled more than 5,000 proteins along the lifespan of the nematode *C. elegans*. We find that one-third of proteins change in abundance at least 2-fold during aging, resulting in a severe proteome imbalance. These changes are reduced in the long-lived *daf-2* mutant but are enhanced in the short-lived *daf-16* mutant. While ribosomal proteins decline and lose normal stoichiometry, proteasome complexes increase. Proteome imbalance is accompanied by widespread protein aggregation, with abundant proteins that exceed solubility contributing most to aggregate load. Notably, the properties by which proteins are selected for aggregation differ in the *daf-2* mutant, and an increased formation of aggregates associated with small heat-shock proteins is observed. We suggest that sequestering proteins into chaperone-enriched aggregates is a protective strategy to slow proteostasis decline during nematode aging.

INTRODUCTION

Protein homeostasis (proteostasis), the state in which the proteome of a living organism is in functional balance, must be tightly controlled within individual cells, tissues, and organs. Maintaining proteome balance requires a complex network of cellular factors, including the machineries of protein synthesis, folding, and degradation (Balch et al., 2008; Hartl et al., 2011), as well as neuronal signaling pathways that regulate proteostasis at the organismal level (Prahald and Morimoto, 2009; Taylor and Dillin, 2013; van Oosten-Hawle and Morimoto, 2014). An important function of these systems is to prevent the accumulation of potentially toxic misfolded and aggregated protein species

(Knowles et al., 2014). However, as organisms age, quality control and the cellular response to unfolded protein stress become compromised (Ben-Zvi et al., 2009; Douglas and Dillin, 2010), and the defense against reactive oxygen species declines (Finkel and Holbrook, 2000). Indeed, aging is considered the principal risk factor for the onset of a number of neurodegenerative disorders associated with aggregate deposition, such as Alzheimer's, Huntington's, and Parkinson's diseases (Knowles et al., 2014). The accumulation of aberrant protein species in these pathologic states in turn places a burden on the proteostasis machinery and thus may accelerate aging by interfering with protein folding and clearance, and other key cellular processes (Balch et al., 2008; Gidalevitz et al., 2006; Hipp et al., 2014; Olzscha et al., 2011). Understanding these relationships requires systematic analyses of the changes that occur in proteome composition and balance during aging.

The nematode *C. elegans* is one of the most extensively studied model organisms in aging research, owing to its relatively short lifespan and the availability of genetic tools to identify pathways that regulate longevity. Inhibition of the insulin/insulin-like growth factor 1 signaling (IIS) pathway in strains carrying mutations in the DAF-2 receptor (or the downstream PI(3) kinase AGE-1) activates the DAF-16/FOXO transcription factor and leads to a dramatic lifespan extension (Kenyon et al., 1993; Murphy et al., 2003). Several lines of evidence suggest that the lifespan-prolonging effect of IIS reduction involves an improvement in cellular stress resistance and proteostasis capacity through upregulation of the machineries mediating protein folding and preventing the formation of toxic aggregate species (Morley et al., 2002; Cohen et al., 2009; Demontis and Perrimon, 2010). In addition to DAF-16 activation, the longevity phenotype in *daf-2* mutants requires the function of HSF-1, the transcription factor regulating the expression of multiple heat-shock proteins and molecular chaperones (Hsu et al., 2003; Morley and Morimoto, 2004). These pathways of proteostasis maintenance appear to be conserved in evolution from worms to mammals (Cohen et al., 2009; Demontis and Perrimon, 2010).

Aging and the effect of the IIS pathway have been studied in *C. elegans* by transcriptome analysis (Budovskaya et al., 2008; Golden and Melov, 2004), but only limited information exists



CrossMark

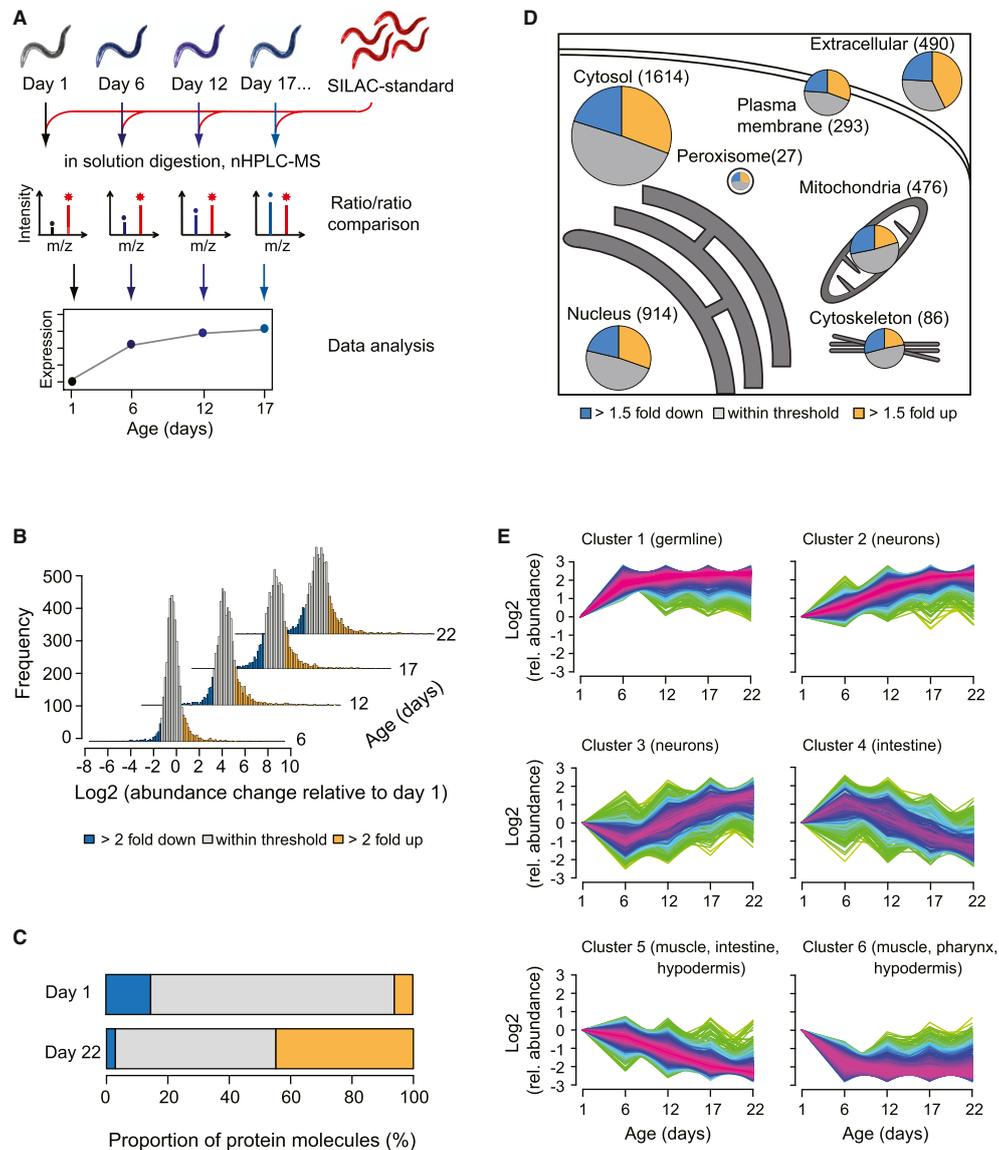


Figure 1. Proteomic Analysis of Aging in *C. elegans*

(A) Experimental design of total proteome analysis. Synchronized worm populations at different time points were lysed and mixed with a metabolically (SILAC) labeled internal protein standard. After digestion, peptides were either analyzed directly or after fractionation by isoelectric focusing, followed by nano-HPLC coupled MS.

(B) Proteome changes in WT animals 6, 12, 17, and 22 days of age relative to day 1 animals (Table S1B). The proportions of proteins that are at least 2-fold increased or decreased in abundance are marked in yellow or blue, respectively.

(C) Contribution to the total proteome of the proteins that change at least 2-fold in abundance between young (day 1) and aged (day 22) animals, as displayed in (B) and estimated by label free quantification (absolute LFQ) (Table S1B).

(D) Proteome changes in subcellular compartments. The fractions of the total proteome that increased (yellow) or decreased (blue) at least 1.5-fold in abundance in old (day 22) versus young (day 1) animals are shown. The color grey represents proteins that remained within the indicated abundance thresholds. Numbers of identified proteins are indicated. Protein subcellular localization was predicted using WoLF PSORT.

(legend continued on next page)

about changes at the proteome level (Dong et al., 2007). Here, we exploit the recent progress in mass spectrometry-based proteomics, which now enables the identification and quantification of thousands of proteins in complex mixtures (Bensimon et al., 2012; Cox and Mann, 2011). We applied stable isotope labeling with amino acids in cell culture (SILAC) (Ong et al., 2002) to profile the abundance levels of more than 5,000 different proteins at multiple time points during the lifespan of *C. elegans*. We then extended our study to short-lived and long-lived strains carrying mutations in the IIS pathway and performed a detailed analysis of age-related protein aggregation. Our data show that during aging, the proteome of the animal undergoes extensive remodeling, escaping proteostasis, and ultimately reaching a state of marked proteome imbalance. These changes are accompanied by widespread protein aggregation, with abundant proteins that exceed their solubility limit making the major contribution to aggregate load. Interestingly, the intrinsic aggregation propensity of proteins is modulated in long-lived *daf-2* mutant worms, resulting in the enhanced formation of chaperone-containing aggregates. Thus, protein aggregation may occur not just as a consequence of proteostasis decline, but may also be induced to improve proteostasis by sequestering surplus, potentially harmful protein species.

RESULTS

Extensive Proteome Remodeling during Aging

To study proteome changes in aging nematodes in depth and with high accuracy, we established a quantitative proteomics approach using SILAC (Ong et al., 2002). Near-complete incorporation of $^{13}\text{C}_6$ - $^{15}\text{N}_2$ -lysine into the proteome was achieved by feeding worms with SILAC labeled ("heavy") *E. coli* cells (Larance et al., 2011). We used a pool of lysates prepared from labeled worms of different ages as internal standards for quantifying protein expression. These standards were added to lysates of synchronized worm populations, followed by digestion and peptide analysis by mass spectrometry (MS) (Figure 1A). Replicate analyses indicated a high degree of reproducibility between individual experiments (Figure S1A; Table S1A). We analyzed the proteomes of adult wild-type (WT) worms from 1 day up to 22 days of age, when less than 30% of the animals remain alive (L4 larval stage defined as day 0). More than 5,000 different proteins were identified and quantified at a false discovery rate of 1% (Table S1B).

Our analysis reveals a broad remodeling of the *C. elegans* proteome during aging. About one-third of the quantified proteins increased or decreased in abundance by at least 2-fold, when equal amounts of total protein were analyzed (Figure 1B; Table S1B). The proteins that increased by at least 2-fold amounted to approximately 50% of total protein in aged animals, as determined by label free absolute quantification (absolute LFQ values) (Schwanhäusser et al., 2011) (Figure 1C). Protein abundance changes were progressive until day 22 (Figures 1B and S1B;

Table S2A) and were observed in most cellular compartments (Figure 1D). Thus, proteome composition and the relative stoichiometries of proteins change dramatically during aging, presumably impeding overall proteostasis. A similar mechanism of proteostasis impairment has been suggested to occur as a result of aneuploidy (Oromendia et al., 2012; Stinglele et al., 2012).

Changes in transcript levels previously observed during aging (Budovskaya et al., 2008; Golden and Melov, 2004) contribute to the changes in protein abundance observed here, but the overall correlation is only moderate ($R = 0.3$) (Figure S1C). Thus, the age-dependent accumulation of a substantial fraction of the proteome is likely to be largely due to posttranscriptional processes. Taking into consideration that microRNA (miRNA)-mediated translational repression of mRNAs is relieved during aging and stress (Ibáñez-Ventoso et al., 2006), we compared our proteome data with a published transcriptome analysis of Dicer mutant worms with defective miRNA biogenesis (Welker et al., 2007). We find that ~30% of proteins that increased more than 2-fold between day 6 and 22 (99 of 357 proteins), i.e., after the worms have reproduced, have significantly elevated transcript levels in dicer mutants, and this proportion increases to nearly 40% for the subset of proteins with a more than 4-fold abundance change (50 out of 133 proteins) (Figure S1D). Thus, miRNA-mediated translational derepression is likely to contribute to the observed increase in protein abundance.

We analyzed the proteomic changes in *C. elegans* aging in terms of various criteria, including subcellular compartments, pathways, and cell types. Among the proteins that increased more than 2-fold in aged worms (22 days) were 183 extracellular proteins (out of 490 extracellular proteins quantified) (Figure S1E; Table S2B). These included multiple transthyretin (TTR)-like factors, which increased up to 100-fold (Figure S1F), as well as all six of the vitellogenin egg storage proteins, despite egg formation having been completed before day 6. Likewise, proteins involved in DNA replication and repair processes were upregulated (Figure S1E), even though all somatic cells of adult *C. elegans* are postmitotic. These examples suggest that many changes in protein abundance during aging do not correlate with biologically relevant activities but instead reflect proteome dysregulation. Among the proteins that declined during aging are nucleolar ribosome biogenesis factors, various peroxisomal enzymes, and proteins involved in lipid glycosylation (Figure S1E; Table S2C). The levels of many mitochondrial proteins also decreased (Figure 1D). For example, subunits of respiratory chain complex I declined gradually by up to 50% during the lifespan (Figure S1G), which may result in the production of reactive oxygen species.

To discern cell-type specific patterns of change, we grouped proteins into clusters using the fuzzy c-means method (Kumar and Futschik, 2007) and analyzed these by tissue-specific expression scores (Chikina et al., 2009) (Figure 1E). We find that age-dependent changes in proteome composition affect

(E) Clustering of time course expression patterns in WT animals using the fuzzy c-means algorithm (Kumar and Futschik, 2007). Significantly enriched tissues as determined by Wilcoxon rank sum test at 2% false discovery rate against predicted expression scores (Chikina et al., 2009) are indicated for each cluster. Warm (red) and cold (blue) colors indicate low and high deviation from the consensus profile, respectively. See also Figure S1 and Tables S1 and S2.

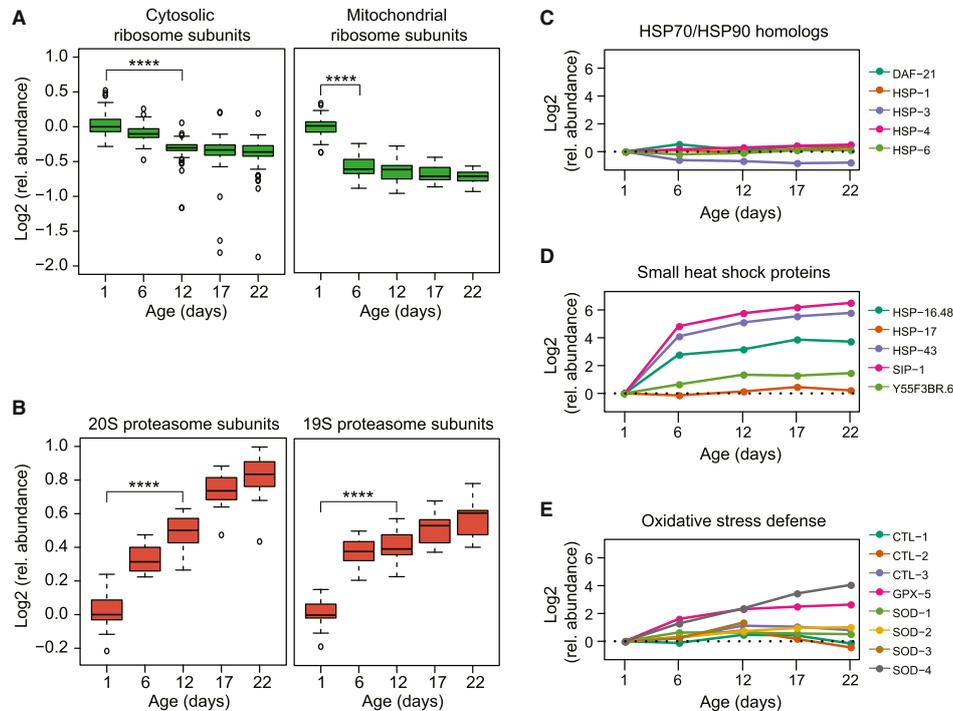


Figure 2. Abundance Changes in Specific Components of the Proteostasis Network

(A) Abundance changes of ribosomal proteins during the lifespan of *C. elegans*. There were 70 different cytosolic (left) and 34 mitochondrial ribosomal proteins (right) that were quantified (see Table S3). Log₂ values of fold-changes are shown in boxplot representation. Solid horizontal lines indicate the median values, whisker caps indicate 10th and 90th percentiles, and circles indicate outliers. ****p < 4.35 × 10⁻¹³ for cytosolic ribosomal proteins and 1.17 × 10⁻¹⁰ for mitochondrial ribosomal proteins from Wilcoxon signed rank test. Only proteins quantified at both time points tested were considered.

(B) Abundance changes of proteasome subunits during lifespan. All 14 subunits of the 20S and 17 subunits of the 19S proteasome were quantified. Only subunits quantified in at least two time points are displayed. ****p < 1.23 × 10⁻⁴ for 20S subunits and ****p < 1.53 × 10⁻⁵ for 19S subunits from Wilcoxon signed rank test. (C–E) Abundance profiles of proteostasis network (PN) components along the lifespan of WT animals. Log₂ relative changes in abundance are shown for HSP70 and HSP90 homologs (C), small HSPs (D), and proteins involved in oxidative stress defense (E). Only components quantified at day 1 and at least three consecutive time points are displayed.

See also Figure S2 and Tables S1 and S3.

a range of tissues. For example, proteins that are predominantly expressed in the germline strongly increase during the first 6 days of adulthood (cluster 1), when the animals reproduce, but surprisingly retain constant levels later in life. Proteins enriched in neuronal cells either increase in abundance throughout the lifespan or after day 6 (clusters 2 and 3). In contrast, the levels of many proteins enriched in intestine, muscle, and hypodermis decline (clusters 4–6), consistent with an age-related deterioration of these tissues.

Age-Related Changes in Proteostasis Network Components

Approximately 440 proteostasis network components involved in protein synthesis, folding, and degradation were quantified throughout the nematode lifespan (Figure S2A; Table S3). A ~25% reduction in the median level of cytosolic ribosomal

proteins occurred between day 1 and day 12 (Figure 2A, left). This reduction correlates with a decrease in the transcript level of ribosome proteins (Golden and Melov, 2004) and an overall age-associated reduction in polysomes (Kirstein-Miles et al., 2013). A similar decrease was observed for mitochondrial ribosomes between day 1 and day 6 (Figure 2A, right). Interestingly, aged animals displayed a pronounced imbalance in the relative subunit stoichiometry of cytosolic, but not mitochondrial, ribosomes, with several subunits decreasing more than 60% below median subunit levels (Figure 2A, left).

Next, we employed SILAC to estimate protein synthesis in aging *C. elegans*. Pulse labeling of worms with heavy bacteria as the food source showed a sharp reduction in the incorporation of labeled amino acids into protein between day 1 and day 4 of adulthood (Figure S2B; Table S1C). This effect was not caused by reduced food uptake, as *eat-2* mutant animals, deficient in

pharyngeal pumping, showed protein labeling equivalent to WT controls, despite their reduced food uptake (data not shown). The reduction in protein synthesis between day 1 and 4 is greater than the decrease in ribosomal levels (Figures 2A and S2B) and probably reflects the reduction in growth of the animals.

In contrast to the effect on ribosomes, we observed an age-dependent increase in 20S and 19S proteasomal subunits (~2-fold at day 22 for 20S subunits) (Figure 2B), correlating with an increase in proteasome activity measured in worm lysates *in vitro* (Figure S2C). Many E3 ubiquitin ligases and other components of the ubiquitin proteasome system (UPS) also increased moderately (Table S3B), while there was no systematic change in the components of autophagy (Figure S2D).

Age-dependent changes in the levels of abundant cytosolic chaperones of the HSP70 and HSP90 (DAF-21) families (Figure 2C) as well as their DnaJ (DNJ/HSP40) and tetratricopeptide repeat (TPR) co-factors were limited (Figures S2E and S2F). Similarly, the subunits of the TRiC/CCT chaperonin remained unchanged (Figure S2G). In contrast, multiple small HSPs, chaperones that function by buffering aggregation, increased dramatically (~13–90-fold), mainly between day 1 and day 6 (Figure 2D). Several of these proteins are under regulation by DAF-16 and HSF-1 (Hsu *et al.*, 2003).

Several components mediating the defense against oxidative stress, including glutathione peroxidase isoform GPX-5 and superoxide dismutases (SOD), increased during aging (up to 12-fold) (Figure 2E; Table S3B). While changes in mitochondrial proteostasis components were generally moderate (Figure S2H; Table S3B), we observed diverse alterations in the proteostasis network of the ER during the nematode lifespan (Figure S2I). For example, protein disulfide isomerases (PDI-2 and C14B9.2), the chaperone calreticulin (CRT-1), as well as the HSP70 homolog HSP-3 decreased ~2-fold, and the pro-collagen modifying enzymes lysyl hydroxylase (LET-268) and prolyl-4-hydroxylase α (DPY-18 and PHY-2) decreased ~3–10-fold. These findings suggest an age-dependent decline in ER quality control and collagen synthesis capacity.

In summary, the levels and activities of two main branches of proteostasis control, protein synthesis and degradation, change in opposite directions during aging. The decrease in ribosomal subunit proteins is accompanied by a dysregulation of cytosolic ribosome assembly, while the increase in proteasome subunits is likely to reflect an attempt at removing surplus or damaged proteins. Other notable changes in the proteostasis system include an increase in the abundance of small HSP chaperones and of components involved in the defense against oxidative stress, as well as a decline in ER protein quality control machinery.

Proteome Changes in Long-Lived and Short-Lived Mutant Strains

To understand in more detail the relationship between the observed proteome changes during the lifespan and the aging process, we next analyzed the proteomes of long-lived *daf-2* (*e1370*), short-lived *daf-16* (*mu86*), and *hsf-1* (*sy441*) mutant worms. The increase in levels of specific proteins observed during aging of WT animals was considerably less pronounced in *daf-2* mutant animals and enhanced in *daf-16* mutant animals

(Figure S3A, left), indicating that the long-lived *daf-2* mutant strain is more effective in controlling the accumulation of surplus proteins. The extent to which proteins decreased in abundance during aging was also greater in *daf-2* mutant worms (Figure S3A, right).

The changes in components of the proteostasis network observed in the mutant strains occurred again predominantly in the protein synthesis and degradation pathways, but at different rates compared to WT. The upregulation of proteasomal subunits commenced earlier during the lifespan of the *daf-2* mutant and was more pronounced than in the WT worms (Figures 3A and 3B); such upregulation was instead less prominent in the short-lived *daf-16* and *hsf-1* mutant strains (Figures 3C and 3D). These results are consistent with the DAF-16 dependent regulation of some proteasome subunits, including RPN6, which is required for 26S proteasome assembly (Vilchez *et al.*, 2012). The decrease in ribosomal proteins occurred at a similar rate in *daf-2* mutant worms as in WT (Figure 3A), but was strongly enhanced in *daf-16* mutant worms (Figure 3C), suggesting that DAF-16 is involved in ribosome maintenance.

Components involved in the oxidative stress response showed marked differences in levels between WT and *daf-2* mutant animals. For example, cytosolic (CTL-1 and CTL-3) and peroxisomal (CTL-2) catalases were 4–8-fold higher in the *daf-2* mutant than in WT worms throughout their lifespans (Figure S3B). SOD-1 (cytoplasmic) and SOD-2 (mitochondrial) were elevated 2-fold compared to WT and short-lived mutant animals (Figure S3C), consistent with their DAF-16-dependent transcriptional regulation (McElwee *et al.*, 2003; Murphy *et al.*, 2003). Among the small HSPs, SIP-1 was already more abundant in young *daf-2* mutant worms (day 1) and HSP-16.48 was markedly elevated in *hsf-1* mutant animals (Figure S3D).

The earlier and more pronounced increase in proteasome abundance in *daf-2* mutant animals may improve the capacity of the organism for the clearance of surplus proteins that accumulate during aging. The elevated levels of catalases and SOD may provide improved defense against oxidative damage.

Age-Dependent Protein Aggregation and Its Relation to Protein Abundance

Declining proteostasis capacity is thought to result in the accumulation of protein aggregates, consistent with recent reports of age-dependent aggregate formation in *C. elegans* (David *et al.*, 2010; Reis-Rodrigues *et al.*, 2012). To analyze this process systematically, we developed a sensitive method for the quantification of aggregated proteins (see Experimental Procedures) and validated it in animals expressing muscle specific FlucDM-GFP, a conformationally unstable mutant of firefly luciferase fused to GFP (Gupta *et al.*, 2011) (Figures S4A and S4B). We isolated insoluble proteins from total lysates of WT animals by centrifugation and performed MS analysis using lysate from labeled worms for quantification (Figure S4C). About 90% of the proteins that were quantified in three out of four experiments (975 of 1,083 proteins) accumulated significantly in the insoluble fraction of day 12 animals relative to day 1 (Table S1D). Age-dependent aggregation was most pronounced between day 6 and day 12 (Figure 4A), i.e., after the hermaphrodite animals ceased to lay eggs. Proteins with predicted transmembrane

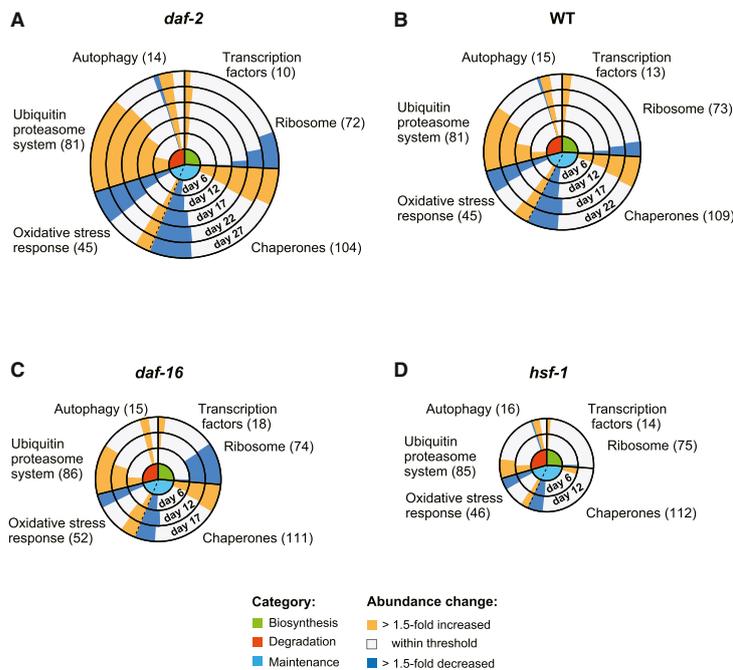


Figure 3. Remodeling of the Proteostasis Network during Aging

(A–D) Abundance changes in components of the PN (see Figure S2A) during aging in *daf-2* (A), WT (B), *daf-16* (C), and *hsf-1* (D) mutant worms. Concentric circles represent increasing age in days from center to periphery. Circle size corresponds with lifespan. Functional categories of components are indicated in the center: green, biosynthesis; red, degradation; and light blue, conformational maintenance (see Figure S2A). Abundance changes of components within these categories relative to day 1 of each strain (yellow, >1.5-fold up and blue, >1.5-fold down) are indicated as bars, with the length of the bar representing the number of proteins undergoing change. The total numbers of proteins quantified in the respective categories are indicated. See also Figure S3.

segments were not enriched in the insoluble fraction (Figure S5A), indicating that lysis was efficient.

To measure the aggregation propensities of proteins during aging, we quantified the insoluble amount of each protein as a fraction of its total amount in aged WT worms (day 12) (Figure S4D; Table S1E). The aggregation propensities of >2,100 analyzed proteins varied by more than two orders of magnitude (Figure 4B), with the median insoluble fraction per individual protein amounting to ~9%.

Previous studies reported a negative correlation between computationally predicted aggregation propensities and protein abundance (Tartaglia et al., 2009). To investigate this dependency at the proteome scale, we grouped proteins according to their aggregation propensities measured at day 12 and estimated the total abundance of each protein in the whole cell lysate by absolute LFQ (Figure 4C). The most abundant proteins were 10-times more soluble than the least abundant proteins. An analysis of the physicochemical properties of the abundant proteins based on their amino acid sequences revealed that they were more hydrophobic (Figure S5B) and more structured (data not shown) than the less abundant ones. These results suggest that abundant proteins increase their solubility, at least in part, by stabilizing their native states through formation of a more extensive hydrophobic core. Indeed, a calculation of the aggregation propensities (Z scores) (Tartaglia et al., 2008; Sormanni et al., 2015a) (see Extended Experimental Procedures) predicts that the more abundant proteins, if correctly folded, are also more soluble (Figure S5C). This conclusion is consistent

with the idea that the solubility of proteins follows their abundance (Tartaglia et al., 2009).

We found, however, that despite of their lower intrinsic aggregation propensities, the most abundant proteins contribute most to the total aggregate load. A strong correlation ($R = 0.75$) was observed between the abundance of specific proteins in the aggregate fraction

and their level in the corresponding whole cell lysate (Figure 4D). Apparently, the high solubility of abundant proteins is insufficient to protect them from age-dependent aggregation, as eventually these proteins exceed their critical concentrations, a phenomenon referred to as “supersaturation” (Ciryam et al., 2013). Notably, we also observed a medium correlation ($R = 0.43$) between the age-dependent change in the total abundance of proteins and their increase in the aggregate fraction (Figure 4E), and this correlation became stronger as aging progressed (data not shown). Thus, proteome remodeling during aging likely drives the aggregation of numerous proteins.

We further investigated whether aggregation also correlates with function. Gene ontology analysis showed that proteins with a relatively high aggregation propensity in aged animals are enriched in the nucleus, whereas abundant glycolytic enzymes and mitochondrial proteins tend to be highly soluble (Figure S5D; Table S4A). Interestingly, all identified small HSPs, but not other chaperones, were highly insoluble at day 12 (Figure 4F), with a high rate of accumulation in the aggregate fraction during aging (Figure S5E). The recruitment of these chaperones into the insoluble fraction may reflect an attempt of the organism to sequester protein aggregates.

Protein Aggregation in Long-Lived and Short-Lived Mutant Strains

Is the age-dependent formation of insoluble aggregates merely a reflection of declining proteostasis capacity, or is it a means to improve proteostasis by sequestering surplus proteins?

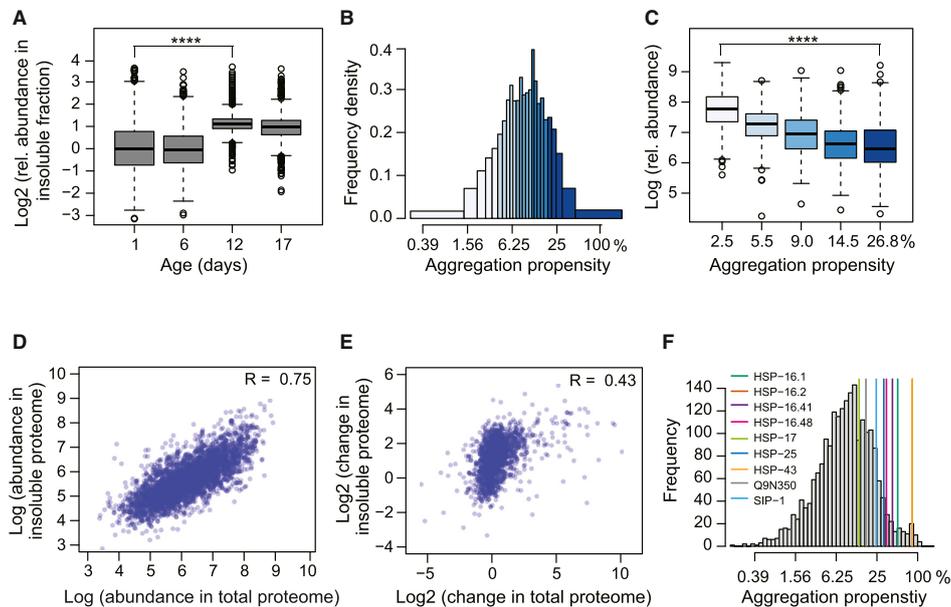


Figure 4. Proteome-wide Analysis of Protein Aggregation during Aging

(A) Relative abundance of proteins in the insoluble fraction of WT animals during aging determined by SILAC quantification (see Figure S4C; Table S1D). At least 1,355 proteins were quantified at the different time points (~3,228 different proteins in total). **** $p < 2.2 \times 10^{-16}$ from Wilcoxon signed rank test.

(B) Distribution of aggregation propensities of proteins (insoluble protein as fraction of total protein) in WT animals at day 12 (median from three independent experiments; Table S1E). Whole worm lysates and insoluble fractions were quantified against the same SILAC standard and ratios were calculated for each protein in % of total (see Figure S4D).

(C) Relationship between aggregation propensity and total protein abundance. Proteins were divided into quantiles based on their measured aggregation propensities (median values are indicated in %). LFQ was used to estimate total protein abundance (displayed as relative abundance values). **** $p < 2.2 \times 10^{-16}$ from Wilcoxon rank sum test.

(D) Protein abundance in the insoluble fraction is positively correlated with abundance in the total proteome (absolute LFQ values). Data for WT animals at day 12 are shown. The Pearson correlation coefficient R is indicated.

(E) Positive correlation between age-related protein abundance changes in the insoluble fraction and abundance changes for the same proteins in the total proteome. Abundance differences measured by SILAC between aged (day 12) and young (day 1) WT animals are plotted. The Pearson correlation coefficient R is indicated.

(F) Aggregation propensities of small HSP family members relative to the aggregation propensities of all quantified proteins in the proteome of day 12 WT animals. See also Figures S4 and S5 and Tables S1 and S4.

Consistent with the former possibility are findings that aggregation-prone model proteins increasingly aggregate in proteostasis-compromised *hsf-1* mutant strains (Ben-Zvi et al., 2009). Indeed, compared to WT animals, the short-lived *hsf-1* mutant worms accumulated more insoluble proteins and aggregation occurred earlier during aging (between day 1 and day 6) (Figures 5A and S6A). However, in support of a beneficial role for aggregation, we found that the long-lived *daf-2* mutant worms also accumulated more insoluble proteins than age-matched WT animals (Figures 5A, S6A, and S6B). This effect was not observed in *daf-16* mutant animals (Figures 5A and S6A), suggesting that age-dependent aggregation is (at least in part) an active process under regulation by DAF-16. The increased aggregation in *daf-2* mutant animals comprised preferentially cytosolic proteins (Figure S6C; Table S4B) and initiated between day 6 and day 12 as in WT (Figure S6A), i.e., when the long-lived mutant worms are still youthful.

While there was a large overlap between the proteins identified in the insoluble fractions, the extent to which specific proteins aggregated varied greatly in a strain specific manner. Interestingly, the proteins that showed increased aggregation in the *daf-2* mutant over WT are not generally more abundant at the total proteome level (Figure 5B), indicating that abundance in this case is not the main driver of aggregation. Similar findings were made in the *hsf-1* mutant (Figure 5C). On the other hand, proteins that aggregated less in the *daf-2* strain than in WT are also generally less abundant (Figure 5B), which would allow these proteins to maintain solubility.

Next, we compared the physico-chemical properties of the insoluble proteins. Strikingly, the proteins that aggregate most in the *daf-2* mutant animals are predicted to have significantly lower aggregation-propensity Z scores, are more charged, display more structural disorder (coil average) (Sormanni et al., 2015b), and are less hydrophobic compared to the proteins

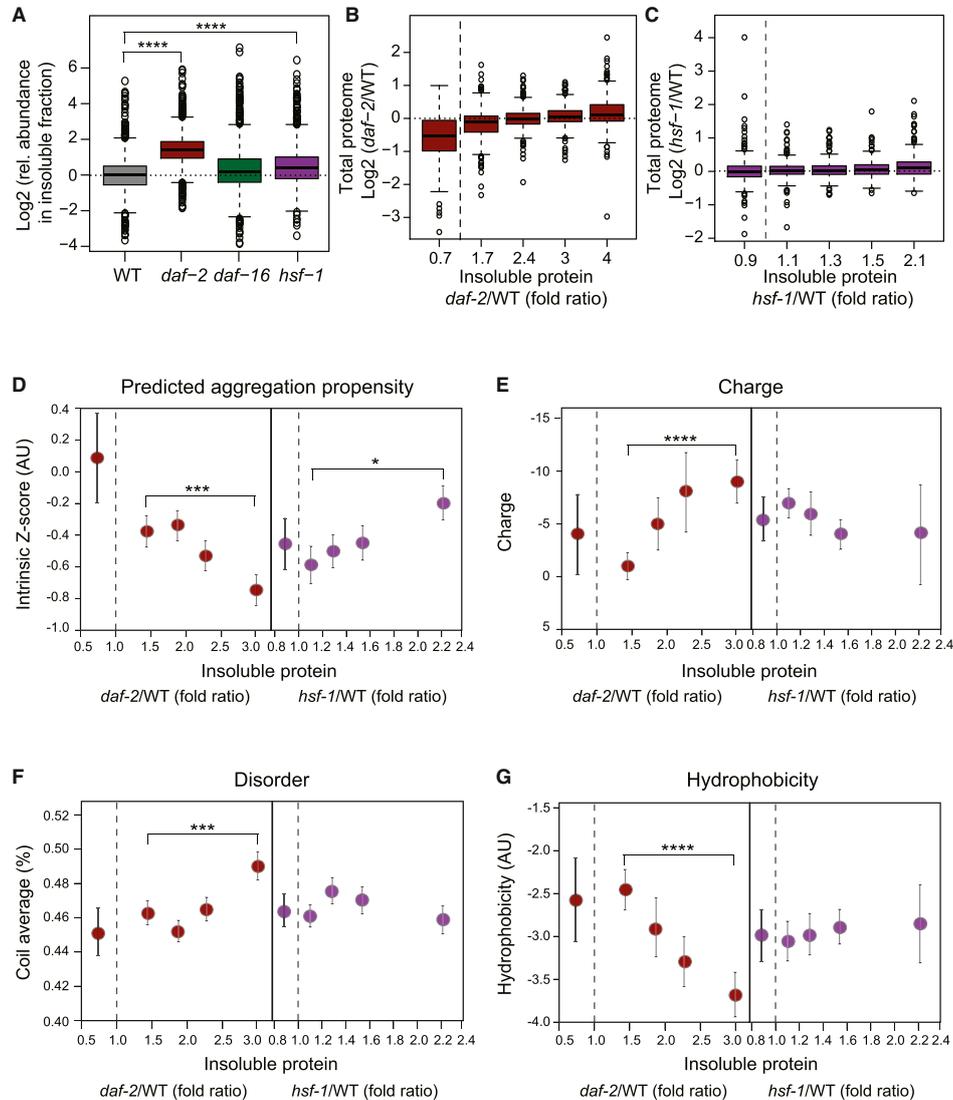


Figure 5. Protein Aggregation in Lifespan Mutant Worms during Aging

(A) Increased aggregate load in *daf-2* mutant animals compared to WT, *daf-16*, and *hsf-1* mutant animals at day 12. Relative abundance values of proteins in the insoluble fraction were determined by SILAC quantification. There were 1,367, 1,988, 1,449, and 1,485 proteins that were quantified in WT, *daf-2*, *daf-16*, and *hsf-1* mutant animals, respectively (one representative out of four independent experiments is displayed; Table S1F). **** $p < 2.2 \times 10^{-16}$ from Wilcoxon signed rank test. (B and C) Quantified abundance of proteins in the insoluble fraction of *daf-2* (352–354 proteins per quantile) (B) and *hsf-1* mutant (292 proteins per quantile) (C) relative to WT animals at day 12 plotted against differences in total protein abundance values. Quantile median values are indicated on the x axis. Proteins that aggregated less in the mutant strains than in the WT have been grouped separately (91 proteins in *daf-2* and 259 in *hsf-1* mutant).

(D–G) Physico-chemical properties of proteins enriched in the insoluble fractions of *daf-2* and *hsf-1* mutant relative to WT animals at day 12.

(D) Aggregation propensity scores (intrinsic Z scores, see Extended Experimental Procedures). *** $p < 1.4 \times 10^{-4}$ and * $p < 0.016$, Wilcoxon rank sum test.

(E) Net charge. **** $p < 4.9 \times 10^{-11}$.

(F) Coil content. *** $p < 1.1 \times 10^{-4}$.

(G) Overall hydrophobicity. **** $p < 2.9 \times 10^{-7}$. Quantile median values are indicated on both axes and standard errors are reported on the y axis.

See also Figure S6 and Table S4.

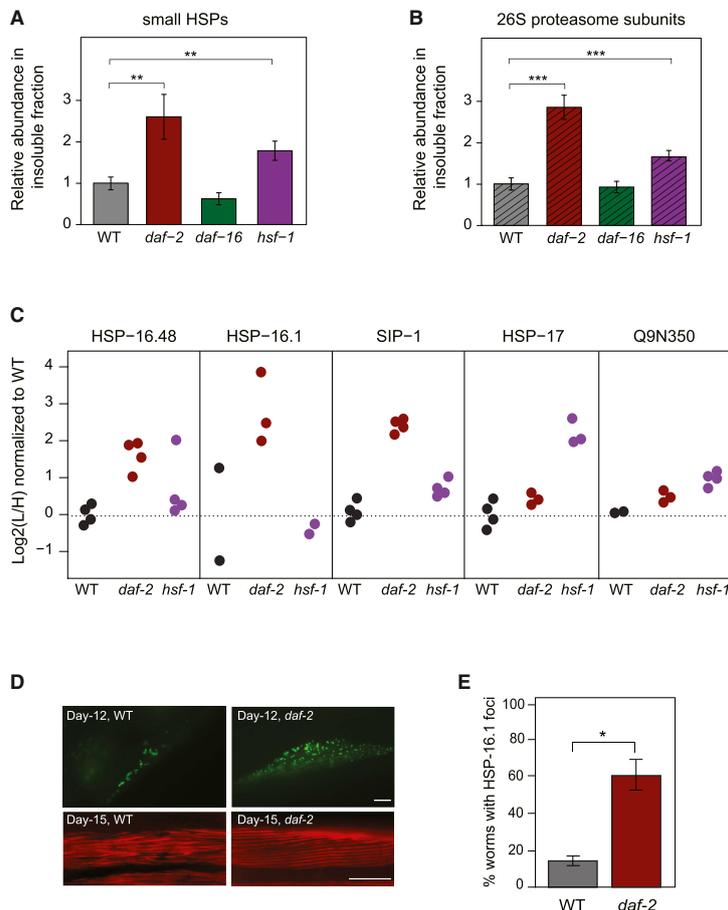


Figure 6. Aggregation of Small HSPs and Proteasome in Lifespan Mutant Worms

(A) Abundance of small HSPs in the insoluble fraction of *daf-2*, *daf-16*, and *hsf-1* mutant relative to WT animals as determined by summed absolute LFQ values. There were 6–11 different small HSPs that were quantified. ***p* value < 0.0075 (WT versus *daf-2*) and < 0.0022 (WT versus *hsf-1*) from Welch's *t* tests. Averages \pm SD are given for four replicate experiments.

(B) Abundance of 26S proteasome subunits in the insoluble fraction of *daf-2*, *daf-16*, and *hsf-1* mutant relative to WT animals. There were 19–27 subunits that were quantified. ****p* < 2.1×10^{-4} (WT versus *daf-2*) and < 4.6×10^{-4} (WT versus *hsf-1*) from Welch's *t* tests. Averages \pm SD are given for four replicate experiments.

(C) Enrichment of the small HSPs HSP-16.1, HSP-16.48, SIP-1, HSP-17, and Q9N350 in the insoluble fractions of day 12 WT (black circles), *daf-2* mutant worms (red circles), and *hsf-1* mutant worms (purple circles). Data from two to four independent experiments are shown.

(D and E) Formation of HSP-16.1 inclusions in muscle cells is shown.

(D) Representative fluorescence images of muscle cells of WT and *daf-2* mutant animals expressing HSP-16.1::GFP (top). Actin was stained with rhodamine-phalloidin (bottom). Scale bar, 10 μ m. (E) Animals with HSP-16.1::GFP inclusions in muscle cells were quantified (20 animals per group). Averages \pm SD are given in % of total. **p* < 0.01 from Welch's *t* test. See also Figure S6 and Table S1.

aggregating in WT (Figures 5D–5G). These findings support the hypothesis of an extrinsic rescuing mechanism of aggregation that is activated in the *daf-2* mutant, modulating the intrinsic properties of proteins that typically govern aggregation. As a result, aggregation is enhanced for a set of proteins that have certain properties resembling disease-associated proteins with structural disorder (Knowles et al., 2014). By contrast, aggregation in the *hsf-1* mutant correlates with intrinsic aggregation scores (Figure 5D), consistent with a degeneration mechanism arising from the premature decline of proteostasis.

Among the proteins that were strongly increased in the insoluble fraction of *daf-2* mutant animals were several small HSPs (Figure 6A). These proteins contribute \sim 7% to total aggregate load, suggesting that they may be involved in a “protective aggregation response”. Small HSPs were also enriched in the insoluble fraction of *hsf-1* mutant animals, although to a lesser extent, but not in the aggregates of the *daf-16* mutant (Figure 6A). Besides small HSPs, 26S proteasome complexes were enriched in the insoluble fractions (Figure 6B), most

made the major contribution by mass to the aggregates in the *daf-2* mutant, while HSP-17 was most enriched in the aggregates of *hsf-1* mutant animals (Figure 6C; Table S1F). To monitor the behavior of HSP-16.1 during aging, we used strains expressing GFP-tagged HSP-16.1 (*hsp-16.1::gfp*) under its endogenous promoter. HSP-16.1-GFP formed inclusions in muscle cells. This phenomenon was strongly enhanced in *daf-2* mutant worms, with \sim 60% of animals at day 12 containing inclusions, compared to \sim 20% in WT (Figures 6D and 6E). While the actin architecture of muscle cells was well preserved in *daf-2* mutant animals, the muscles of day 15 WT animals showed a reduced structural integrity (Figure 6D). Indeed, the *daf-2* mutant animals displayed a higher proteostasis capacity, as reflected in their ability to maintain the metastable FlucDM-GFP (Gupta et al., 2011) expressed in muscle in a functionally active state. While similar levels of total and soluble FlucDM-GFP protein were present in day 12 WT and *daf-2* mutant worms, the latter contained \sim 4-fold more luciferase activity (Figure S6D). In contrast, a muscle-specific poly-glutamine (polyQ) protein construct (Q35-GFP)

strongly in the *daf-2* mutant strain, but contributed only \sim 1% to total aggregate load.

Interestingly, the proportion of specific small HSPs in the aggregates differed between strains. SIP-1 and HSP-16.1

aggregated more extensively in *daf-2* mutant worms already early in adulthood (day 2), and semi-denaturing detergent agarose gel electrophoresis (SDD-AGE) of worm lysates revealed that the protein accumulated predominantly in higher molecular weight, SDS-resistant oligomers (Figure S6E).

Taken together, these results suggest that *daf-2* mutant animals drive a set of aberrant, potentially toxic proteins into insoluble aggregates, thereby sequestering them and improving overall proteostasis. Small HSPs are likely to play a role in this process.

DISCUSSION

Age-Dependent Deterioration of Proteome Balance

Organisms allocate considerable resources toward maintaining proteome composition, including the relative balance of subunits of multi-protein complexes (Li et al., 2014). Using quantitative mass spectrometric methods, we have shown here that aging in *C. elegans* is associated with the progressive failure to maintain protein homeostasis, resulting in extensive proteome remodeling and protein imbalances. These imbalances are largely due to changes at the level of protein translation and turnover and give rise to the accumulation of potentially harmful, aggregation-prone species (Figures 7A and 7B). Our analysis revealed that the sequestration of such proteins in insoluble aggregates is a protective strategy that contributes to maintaining proteome integrity during aging.

The extensive proteome remodeling during aging in *C. elegans* is contrary to observations in tissues of aged mice, where comparatively minor proteomic changes were detected with a similar experimental approach (Walther and Mann, 2011). Evidently, mammals devote greater resources to maintaining proteome balance, resulting in a more protracted proteostasis decline. These differences correlate with different reproductive strategies in worms and mice, in which the former display a more extensive, time-restrained burst of reproduction, followed by a rather rapid and massive decline of somatic functions. Future studies on a range of metazoans will be necessary to establish whether deterioration of proteome integrity during aging or proteome stability is more typical.

Changes in the Proteostasis System during Lifespan

We showed that aging in *C. elegans* affects multiple components of the proteostasis system, most prominently protein biosynthesis and protein degradation. A decrease in the levels of cytosolic and mitochondrial ribosomal proteins was accompanied by an overall reduction in protein synthesis. In contrast, we observed an increase in the abundance of proteasome subunits and a corresponding increase of *in vitro* proteasome activity. These changes may initiate as a response to the altered physiological requirements of the aging organism (Shore and Ruvkun, 2013), but ultimately may prove insufficient or even detrimental (Figure 7B). The reduction of the levels of cytosolic ribosomes was associated with a pronounced imbalance in the stoichiometry of ribosomal proteins. Thus, attenuating the translational machinery as an adaptive measure imparts the danger of dysregulation of the essential machines that ensure balance of the proteostasis network, which in turn may promote

aging. In contrast, the increase in proteasomal subunits is likely to represent an attempt of the organism to remove aberrant protein species. Whether this proteasome upregulation is effective *in vivo* is unclear, however, given that proteasome function is generally thought to decline as a result of aging and protein aggregation (Hipp et al., 2014).

Protein Aggregation and Lifespan Extension

Previous studies demonstrated the formation of insoluble protein aggregates in aged worms (David et al., 2010; Reis-Rodrigues et al., 2012). Here, we performed an in-depth quantitative analysis of aggregation along the lifespan of *C. elegans*. We found that aggregation is a proteome-wide process which initiates mainly after day 6 of adulthood. Highly abundant proteins are generally more soluble and display lower intrinsic aggregation-propensities than less abundant ones, as previously predicted (Tartaglia et al., 2009). However, this higher solubility is still not sufficient in the end, as abundant proteins make by far the major contribution by mass to the age-dependent aggregate load. Importantly, proteome remodeling acts as a driver of aggregation by raising the level of a subset of proteins beyond a critical solubility limit (supersaturation) (Ciryam et al., 2013) (Figure 7D).

While protein aggregation may be merely a consequence of declining proteostasis capacity, our results provide evidence that a protective aggregation response is also an important mechanism of the aging organism to improve proteostasis and mitigate the effects of proteome imbalance. We observed that long-lived *daf-2* mutant animals accumulate increasing amounts of insoluble proteins during aging and that such accumulation correlates with a more effective maintenance of proteome composition (Figure 7C). Whereas the proteins that aggregate most in the short-lived *hsf-1* mutant are predicted to be more aggregation-prone, the enhanced aggregation in the long-lived *daf-2* mutant is much less dependent on intrinsic properties: the proteins that are most enriched in the insoluble fraction have lower aggregation scores, are less hydrophobic, more charged, and contain more structural disorder, arguing for the existence of an active, proteome-wide mechanism in promoting aggregation. This conclusion is consistent with the view that soluble oligomers are the major proteotoxic species in neurodegenerative diseases and that their sequestration into insoluble aggregates reduces proteotoxicity (Arrasate et al., 2004; Cohen et al., 2006, 2009). Interestingly, several highly toxic disease-associated proteins are rich in disordered structure and have low overall hydrophobicity (Knowles et al., 2014; Vendruscolo et al., 2011), properties resembling those of the proteins with enhanced aggregation in the *daf-2* mutant. Indeed, a mechanism of aggregate deposition under regulation of the insulin signaling pathway has been proposed for disease-related protein species, such as toxic A β peptide (Cohen et al., 2006). However, that an overall protective aggregate response operates at the proteome-scale during aging was not anticipated.

We assume that this protective aggregation response is only partially activated during normal *C. elegans* aging. As a result, WT worms are expected to accumulate a larger soluble pool of aberrant, potentially toxic proteins than *daf-2* mutant animals, eventually exhausting the available chaperone capacity needed for protein folding and conformational maintenance and the

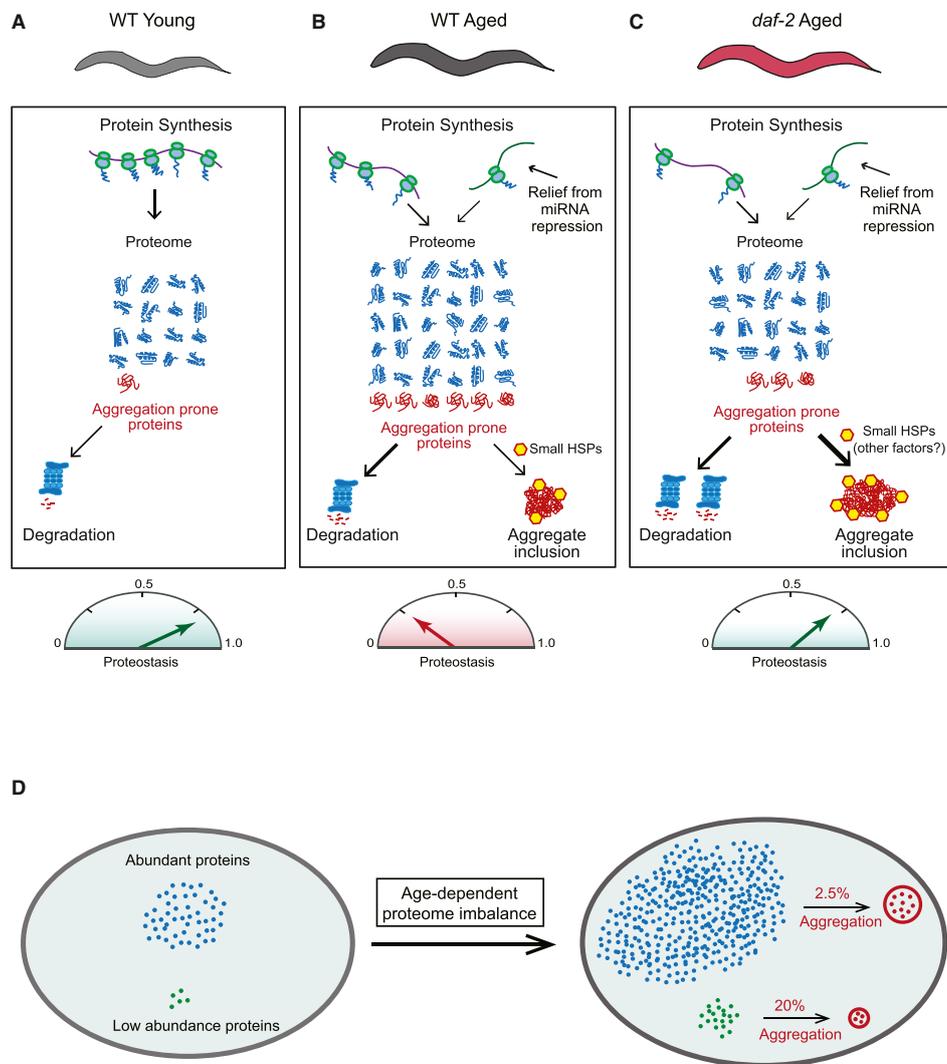


Figure 7. Proteome Maintenance during Aging in *C. elegans*

(A) The proteome of young adult WT worms is maintained in balance by the proteostasis system. Aberrant protein species, including metastable conformers and soluble aggregates (red) are efficiently cleared.

(B) In aged WT animals, numerous proteins increase in abundance and normal protein stoichiometries are lost due in part to a relief of miRNA-mediated translational repression. The amount of aggregation-prone species exceeds clearance capacity and insoluble aggregates associated with small HSPs accumulate. Mechanisms of protective aggregate formation are partially activated, and proteostasis is strongly reduced.

(C) Proteostasis collapse is delayed in aged *daf-2* mutant worms. Proteome imbalance and the soluble aggregate pool are reduced relative to age-matched WT animals, as clearance by protein degradation may be more effective and protective aggregate formation is fully activated.

(D) Protein aggregate loads increase proportionally to protein abundance. Although abundant proteins have lower aggregation propensities, they contribute more to aggregate load (see Figure 4). The age-dependent increase in expression level affects the subproteome of supersaturated proteins, which fail to maintain solubility as their levels increase and proteostasis capacity declines.

clearance of misfolded polypeptides (Figures 7B and 7C). Formation of insoluble aggregates may also be an, albeit insufficient, rescue attempt in the short-lived *hsf-1* mutant strain.

Among the proteostasis components with strongly enhanced, age-dependent insolubility were multiple small HSPs, a specific class of chaperones known to associate with aggregation-prone

proteins (Haslbeck et al., 2005). The small HSPs were most enriched in the insoluble fraction of *daf-2* mutant worms, consistent with the view that they may play a role as “extrinsic” promoters of aggregation. In support of this possibility, individual RNAi knockdown of several small HSPs, including SIP-1, caused a 25% shortening of lifespan in WT and *daf-2* mutant worms (Hsu et al., 2003) and overexpression resulted in lifespan extension (Walker and Lithgow, 2003). Having multi-valent binding ability for aberrant proteins, the small HSPs may act to seed and concentrate aggregate material, consistent with findings in vivo (Escusa-Toret et al., 2013; Kaganovich et al., 2008; Specht et al., 2011) and in vitro (Haslbeck et al., 2005; Jiao et al., 2005). The co-existence of multiple small HSPs suggests that different forms may vary in their structural specificity for endogenous proteins. Notably, small HSPs are also transcriptionally induced in the aging brain, while most other major chaperone components are downregulated (Brehme et al., 2014). The association of the 26S proteasome with the aggregates may also be functionally relevant. Although aggregates can inhibit the proteolytic activity of the proteasome (Andersson et al., 2013; Hipp et al., 2014), evidence has been presented that the ATPase chaperones of the 19S proteasome may promote aggregation (Rousseau et al., 2009).

Collectively, our data suggest that aging in *C. elegans* is associated with a progressive loss of proteome balance, which drives the accumulation of surplus and aberrant protein species that overburden the proteostasis system. As the maintenance of protein solubility imposes stringent constraints on proteome composition, effective aggregate management appears to be critical in determining lifespan.

EXPERIMENTAL PROCEDURES

C. elegans Strains and Growth Conditions

A list of strains used in this study is provided in the Extended Experimental Procedures. The Bristol strain N2 was used as WT. The L4 larval stage was considered as day 0. Bacterial cultures (ET505) for SILAC labeling were grown in ¹³C₆-¹⁵N₂-lysine (heavy lysine) containing M63 minimal media (Krijgsveld et al., 2003) (see Extended Experimental Procedures for details).

Sample Preparation for Total Proteome Measurements

Briefly, worms were suspended in lysis buffer (4% SDS, 0.1 M Tris/HCl pH 8.0, and 1 mM EDTA), incubated at 95°C for 5 min, and further treated by ultrasonication. Typically, an aliquot of lysate containing 40 μg of protein was mixed with an equal amount of a heavy lysine labeled lysate pool (Figure 1A). Proteins were reduced, alkylated, and digested with endoproteinase LysC using the filter-aided sample preparation (FASP) method (Wiśniewski et al., 2009). Peptide mixtures were either analyzed without fractionation or after fractionation by isoelectric focusing, as described in the Extended Experimental Procedures.

Isolation of Protein Aggregates

Worms were resuspended in lysis buffer (50 mM Tris/HCl pH 8.0, 0.5 M NaCl, 4 mM EDTA, 1% volume/volume (v/v) Igepal CA630, and complete protease inhibitor cocktail; Roche Diagnostics), disrupted by sonication, and clarified by low-speed centrifugation (1 min, 1,000 relative centrifugal force [rcf]). Insoluble proteins were sedimented by ultracentrifugation (500,000 rcf at 4°C, 10 min), washed twice with lysis buffer containing 0.15 M NaCl and 0.5% sodium deoxycholate, and solubilized in SDS sample buffer for 10 min at 95°C. For quantitative proteome measurements of aggregated proteins, an aliquot of pooled total lysate from heavy lysine labeled animals was added prior to ultracentrifugation. For experiments measuring aggregation propen-

sities, unlabeled worm lysates were first fractionated and subsequently supplemented with SILAC-labeled whole cell lysate.

MS and Data Analysis

Peptides were separated by reversed phase nano-high-performance liquid chromatography (HPLC) and sprayed online into LTQ-Orbitrap Velos or Orbitrap-Elite mass spectrometers (Thermo Fisher). In each scan cycle, fragmentation spectra of the ten most intense peptide precursors in the survey scan were acquired in the higher-energy collisional dissociation (HCD) mode. Raw data were processed using the MaxQuant software environment (Cox and Mann, 2008) and peak lists were searched with Andromeda (Cox et al., 2011) against a database containing the translation of all predicted proteins listed in UniProt (release January 15, 2012), as well as a list of commonly observed contaminants and the National Center for Biotechnology Information (NCBI) protein database of *E. coli* strain K12. The minimal required peptide length was set to seven amino acids and both protein and peptide identifications were accepted at a false discovery rate of 1%. To identify aggregation-prone proteins that were significantly affected by aging, those proteins that were quantified in at least three out of four biological replicate experiments at day 1 and day 12 were subjected to a Welch's t test and filtered based on a 5% permutation-based false discovery rate threshold.

Miscellaneous

Proteasome activity assays, detection of polyQ aggregates by SDD-AGE, light microscopy, and methods used for bioinformatic analyses are described in the Extended Experimental Procedures.

ACCESSION NUMBERS

Proteomics raw data and selected MaxQuant output files have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD001364.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and four tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.03.032>.

AUTHOR CONTRIBUTIONS

F.U.H., M.M., D.M.W., and P.K. conceived the approach. D.M.W. and P.K. designed and performed the experiments with contributions from M.Z. M.M. supervised the proteomics analyses. D.M.W. and S.P. performed bioinformatics analyses with contributions from P.K. G.V. and P.C. analyzed the protein aggregation data and interpreted the results together with M.V., C.M.D., and R.I.M. F.U.H., D.M.W., and P.K. wrote the manuscript with contributions from M.M., M.V., C.M.D., and R.I.M.

ACKNOWLEDGMENTS

We thank the Caenorhabditis Genetics Center for providing most of the strains used in this study. *hsp-16.1::gfp* transgenic worms were provided by Dr. Junho Lee, Seoul National University. This work was supported by the European Commission under FP7 GA number ERC-2012-SyG_318987-Toxic Protein Aggregation in Neurodegeneration (ToPAG), the Wellcome Trust (094425/Z/10/Z), the Center for Integrated Protein Science Munich (CIPSM), and the Munich Cluster for Systems Neurology (SyNergy). We would further like to thank Igor Paron and Korbinian Mayr for excellent technical assistance with MS instrumentation, as well as Daniel Hornburg, Manajit Hayer-Hartl, and members of the Hartl laboratory for critically reading this manuscript.

Received: October 23, 2014

Revised: January 12, 2015

Accepted: February 24, 2015

Published: May 7, 2015

REFERENCES

- Andersson, V., Hanzén, S., Liu, B., Molin, M., and Nyström, T. (2013). Enhancing protein disaggregation restores proteasome activity in aged cells. *Aging (Albany, N.Y. Online)* 5, 802–812.
- Arrasate, M., Mitra, S., Schweitzer, E.S., Segal, M.R., and Finkbeiner, S. (2004). Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature* 431, 805–810.
- Balch, W.E., Morimoto, R.I., Dillin, A., and Kelly, J.W. (2008). Adapting proteostasis for disease intervention. *Science* 319, 916–919.
- Ben-Zvi, A., Miller, E.A., and Morimoto, R.I. (2009). Collapse of proteostasis represents an early molecular event in *Caenorhabditis elegans* aging. *Proc. Natl. Acad. Sci. USA* 106, 14914–14919.
- Bensimon, A., Heck, A.J., and Aebersold, R. (2012). Mass spectrometry-based proteomics and network biology. *Annu. Rev. Biochem.* 81, 379–405.
- Brehme, M., Voisine, C., Rolland, T., Wachi, S., Soper, J.H., Zhu, Y., Orton, K., Vilella, A., Garza, D., Vidal, M., et al. (2014). A chaperone subnetwork safeguards proteostasis in aging and neurodegenerative disease. *Cell Rep.* 9, 1135–1150.
- Budovskaya, Y.V., Wu, K., Southworth, L.K., Jiang, M., Tedesco, P., Johnson, T.E., and Kim, S.K. (2008). An elt-3/elt-5/elt-6 GATA transcription circuit guides aging in *C. elegans*. *Cell* 134, 291–303.
- Chikina, M.D., Huttenhower, C., Murphy, C.T., and Troyanskaya, O.G. (2009). Global prediction of tissue-specific gene expression and context-dependent gene networks in *Caenorhabditis elegans*. *PLoS Comput. Biol.* 5, e1000417.
- Ciryam, P., Tartaglia, G.G., Morimoto, R.I., Dobson, C.M., and Vendruscolo, M. (2013). Widespread aggregation and neurodegenerative diseases are associated with supersaturated proteins. *Cell Rep.* 5, 781–790.
- Cohen, E., Bieschke, J., Perciavalle, R.M., Kelly, J.W., and Dillin, A. (2006). Opposing activities protect against age-onset proteotoxicity. *Science* 313, 1604–1610.
- Cohen, E., Paulsson, J.F., Blinder, P., Burstyn-Cohen, T., Du, D., Estepa, G., Adame, A., Pham, H.M., Holzenberger, M., Kelly, J.W., et al. (2009). Reduced IGF-1 signaling delays age-associated proteotoxicity in mice. *Cell* 139, 1157–1169.
- Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367–1372.
- Cox, J., and Mann, M. (2011). Quantitative, high-resolution proteomics for data-driven systems biology. *Annu. Rev. Biochem.* 80, 273–299.
- Cox, J., Neuhauser, N., Michalski, A., Scheiterna, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* 10, 1794–1805.
- David, D.C., Ollikainen, N., Trinidad, J.C., Cary, M.P., Burlingame, A.L., and Kenyon, C. (2010). Widespread protein aggregation as an inherent part of aging in *C. elegans*. *PLoS Biol.* 8, e1000450.
- Demontis, F., and Perrimon, N. (2010). FOXO/4E-BP signaling in *Drosophila* muscles regulates organism-wide proteostasis during aging. *Cell* 143, 813–825.
- Dong, M.Q., Venable, J.D., Au, N., Xu, T., Park, S.K., Cociorva, D., Johnson, J.R., Dillin, A., and Yates, J.R., 3rd. (2007). Quantitative mass spectrometry identifies insulin signaling targets in *C. elegans*. *Science* 317, 660–663.
- Douglas, P.M., and Dillin, A. (2010). Protein homeostasis and aging in neurodegeneration. *J. Cell Biol.* 190, 719–729.
- Escusa-Toret, S., Vonk, W.I., and Frydman, J. (2013). Spatial sequestration of misfolded proteins by a dynamic chaperone pathway enhances cellular fitness during stress. *Nat. Cell Biol.* 15, 1231–1243.
- Finkel, T., and Holbrook, N.J. (2000). Oxidants, oxidative stress and the biology of ageing. *Nature* 408, 239–247.
- Gidalevitz, T., Ben-Zvi, A., Ho, K.H., Brignull, H.R., and Morimoto, R.I. (2006). Progressive disruption of cellular protein folding in models of polyglutamine diseases. *Science* 311, 1471–1474.
- Golden, T.R., and Melov, S. (2004). Microarray analysis of gene expression with age in individual nematodes. *Aging Cell* 3, 111–124.
- Gupta, R., Kasturi, P., Bracher, A., Loew, C., Zheng, M., Vilella, A., Garza, D., Hartl, F.U., and Raychaudhuri, S. (2011). Firefly luciferase mutants as sensors of proteome stress. *Nat. Methods* 8, 879–884.
- Hartl, F.U., Bracher, A., and Hayer-Hartl, M. (2011). Molecular chaperones in protein folding and proteostasis. *Nature* 475, 324–332.
- Haslbeck, M., Franzmann, T., Weinfurter, D., and Buchner, J. (2005). Some like it hot: the structure and function of small heat-shock proteins. *Nat. Struct. Mol. Biol.* 12, 842–846.
- Hipp, M.S., Park, S.H., and Hartl, F.U. (2014). Proteostasis impairment in protein-misfolding and -aggregation diseases. *Trends Cell Biol.* 24, 506–514.
- Hsu, A.-L., Murphy, C.T., and Kenyon, C. (2003). Regulation of aging and age-related disease by DAF-16 and heat-shock factor. *Science* 300, 1142–1145.
- Ibáñez-Ventoso, C., Yang, M., Guo, S., Robins, H., Padgett, R.W., and Driscoll, M. (2006). Modulated microRNA expression during adult lifespan in *Caenorhabditis elegans*. *Aging Cell* 5, 235–246.
- Jiao, W., Li, P., Zhang, J., Zhang, H., and Chang, Z. (2005). Small heat-shock proteins function in the insoluble protein complex. *Biochem. Biophys. Res. Commun.* 335, 227–231.
- Kaganovich, D., Kopito, R., and Frydman, J. (2008). Misfolded proteins partition between two distinct quality control compartments. *Nature* 454, 1088–1095.
- Kenyon, C., Chang, J., Gensch, E., Rudner, A., and Tabtiang, R. (1993). A *C. elegans* mutant that lives twice as long as wild type. *Nature* 366, 461–464.
- Kirstein-Miles, J., Scior, A., Deuring, E., and Morimoto, R.I. (2013). The nascent polypeptide-associated complex is a key regulator of proteostasis. *EMBO J.* 32, 1451–1468.
- Knowles, T.P., Vendruscolo, M., and Dobson, C.M. (2014). The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.* 15, 384–396.
- Krijgsveld, J., Ketting, R.F., Mahmoudi, T., Johansen, J., Artal-Sanz, M., Verrijzer, C.P., Plasterk, R.H.A., and Heck, A.J.R. (2003). Metabolic labeling of *C. elegans* and *D. melanogaster* for quantitative proteomics. *Nat. Biotechnol.* 21, 927–931.
- Kumar, L., and Futschik, M. (2007). Mfuzz: a software package for soft clustering of microarray data. *Bioinformatics* 2, 5–7.
- Larance, M., Bailly, A.P., Pourkarimi, E., Hay, R.T., Buchanan, G., Coulthurst, S., Xirodimas, D.P., Gartner, A., and Lamond, A.I. (2011). Stable-isotope labeling with amino acids in nematodes. *Nat. Methods* 8, 849–851.
- Li, G.W., Burkhardt, D., Gross, C., and Weissman, J.S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* 157, 624–635.
- McEwee, J., Bubba, K., and Thomas, J.H. (2003). Transcriptional outputs of the *Caenorhabditis elegans* forkhead protein DAF-16. *Aging Cell* 2, 111–121.
- Morley, J.F., and Morimoto, R.I. (2004). Regulation of longevity in *Caenorhabditis elegans* by heat shock factor and molecular chaperones. *Mol. Biol. Cell* 15, 657–664.
- Morley, J.F., Brignull, H.R., Weyers, J.J., and Morimoto, R.I. (2002). The threshold for polyglutamine-expansion protein aggregation and cellular toxicity is dynamic and influenced by aging in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* 99, 10417–10422.
- Murphy, C.T., McCarroll, S.A., Bargmann, C.I., Fraser, A., Kamath, R.S., Ahringer, J., Li, H., and Kenyon, C. (2003). Genes that act downstream of DAF-16 to influence the lifespan of *Caenorhabditis elegans*. *Nature* 424, 277–283.
- Olzsha, H., Schermann, S.M., Woerner, A.C., Pinkert, S., Hecht, M.H., Tartaglia, G.G., Vendruscolo, M., Hayer-Hartl, M., Hartl, F.U., and Vabulas, R.M. (2011). Amyloid-like aggregates sequester numerous metastable proteins with essential cellular functions. *Cell* 144, 67–78.
- Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture,

- SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* 7, 376–386.
- Oromendia, A.B., Dodgson, S.E., and Amon, A. (2012). Aneuploidy causes proteotoxic stress in yeast. *Genes Dev.* 26, 2696–2708.
- Prahlad, V., and Morimoto, R.I. (2009). Integrating the stress response: lessons for neurodegenerative diseases from *C. elegans*. *Trends Cell Biol.* 19, 52–61.
- Reis-Rodrigues, P., Czerwieńiec, G., Peters, T.W., Evani, U.S., Alavez, S., Gaman, E.A., Vantipalli, M., Mooney, S.D., Gibson, B.W., Lithgow, G.J., and Hughes, R.E. (2012). Proteomic analysis of age-dependent changes in protein solubility identifies genes that modulate lifespan. *Aging Cell* 11, 120–127.
- Rousseau, E., Kojima, R., Hoffner, G., Djian, P., and Bertolotti, A. (2009). Misfolding of proteins with a polyglutamine expansion is facilitated by proteasomal chaperones. *J. Biol. Chem.* 284, 1917–1929.
- Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature* 473, 337–342.
- Shore, D.E., and Ruvkun, G. (2013). A cytoprotective perspective on longevity regulation. *Trends Cell Biol.* 23, 409–420.
- Sormanni, P., Aprile, F.A., and Vendruscolo, M. (2015a). The CamSol method of rational design of protein mutants with enhanced solubility. *J. Mol. Biol.* 427, 478–490.
- Sormanni, P., Camilloni, C., Fariselli, P., and Vendruscolo, M. (2015b). The s2D method: simultaneous sequence-based prediction of the statistical populations of ordered and disordered regions in proteins. *J. Mol. Biol.* 427, 982–996.
- Specht, S., Miller, S.B., Mogk, A., and Bukau, B. (2011). Hsp42 is required for sequestration of protein aggregates into deposition sites in *Saccharomyces cerevisiae*. *J. Cell Biol.* 195, 617–629.
- Stingele, S., Stoehr, G., Peplowska, K., Cox, J., Mann, M., and Storchova, Z. (2012). Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells. *Mol. Syst. Biol.* 8, 608.
- Tartaglia, G.G., Pawar, A.P., Campioni, S., Dobson, C.M., Chiti, F., and Vendruscolo, M. (2008). Prediction of aggregation-prone regions in structured proteins. *J. Mol. Biol.* 380, 425–436.
- Tartaglia, G.G., Pechmann, S., Dobson, C.M., and Vendruscolo, M. (2009). A relationship between mRNA expression levels and protein solubility in *E. coli*. *J. Mol. Biol.* 388, 381–389.
- Taylor, R.C., and Dillin, A. (2013). XBP-1 is a cell-nonautonomous regulator of stress resistance and longevity. *Cell* 153, 1435–1447.
- van Oosten-Hawle, P., and Morimoto, R.I. (2014). Organismal proteostasis: role of cell-nonautonomous regulation and transcellular chaperone signaling. *Genes Dev.* 28, 1533–1543.
- Vendruscolo, M., Knowles, T.P., and Dobson, C.M. (2011). Protein solubility and protein homeostasis: a generic view of protein misfolding disorders. *Cold Spring Harb. Perspect. Biol.* 3, a010454.
- Vilchez, D., Morante, I., Liu, Z., Douglas, P.M., Merkwirth, C., Rodrigues, A.P., Manning, G., and Dillin, A. (2012). RPN-6 determines *C. elegans* longevity under proteotoxic stress conditions. *Nature* 489, 263–268.
- Walker, G.A., and Lithgow, G.J. (2003). Lifespan extension in *C. elegans* by a molecular chaperone dependent upon insulin-like signals. *Aging Cell* 2, 131–139.
- Walther, D.M., and Mann, M. (2011). Accurate quantification of more than 4000 mouse tissue proteins reveals minimal proteome changes during aging. *Mol. Cell. Proteomics* 10, M110.004523.
- Welker, N.C., Habig, J.W., and Bass, B.L. (2007). Genes misregulated in *C. elegans* deficient in Dicer, RDE-4, or RDE-1 are enriched for innate immunity genes. *RNA* 13, 1090–1102.
- Wiśniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nat. Methods* 6, 359–362.

Supplemental Information

Cell

EXTENDED EXPERIMENTAL PROCEDURES

Strains

C. elegans strains were maintained by standard methods (Brenner, 1974). The Bristol strain N2 was used as wild-type. The following mutants and transgenic strains were used: CB1370 [*daf-2 (e1370)III*], CF1038 [*daf-16 (mu86)I*], PS3551 [*hsf-1 (sy441)I*], DA1113 [*eat-2 (ad1113)II*], AM140 [*rmls132 [P(unc-54) q35::yfp]*], FUH135 [*marls135 [P(unc-54) Fluc-DM::gfp]+rol-6 (su1006)*], FUH139 [*marls139 [P(unc-54) Fluc-DM::gfp]+rol-6 (su1006)*], FUH179 [*daf-2 (e1370); marls139 [P(unc-54)Fluc-DM::gfp]+rol-6 (su1006)*], FUH219 [*daf-2 (e1370); rmls132 [P(unc-54) q35::yfp]*], FUH236 [*P(hsp-16.1) hsp-16.1::gfp]+rol-6 (su1006)*] and FUH237 [*daf-2 (e1370); P(hsp-16.1) hsp-16.1::gfp]+rol-6 (su1006)*].

Growth Conditions

Bacterial cultures (ET505) for SILAC labeling were grown in $^{13}\text{C}_6$ - $^{15}\text{N}_2$ -lysine (heavy lysine) containing M63 minimal media, harvested by centrifugation and washed. Suspensions were spotted onto nitrogen-free agarose plates (Krijgsveld et al., 2003). The incorporation of heavy lysine into the proteome was more than 99% after four reproductive cycles. Worm eggs were collected and synchronized populations of L1 larvae were obtained by overnight growth in M9 medium. The L4 larval stage was considered to be day 0 and larvae were transferred at this time point to new plates with or without 10 μM fluorodeoxyuridine (FUdR). For total proteome measurements FUdR was omitted and progeny was removed by repeated sedimentation. Dead worms were removed manually before harvesting.

Sample Preparation for Total Proteome Measurements

Worms were rinsed off plates and washed with M9 salt solution to minimize bacterial contamination. Worm pellets were resuspended in lysis buffer (4% SDS, 0.1 M Tris/HCl pH 8.0, 1 mM EDTA), incubated at 95°C for 5 min and sonicated in a Bioruptor (Diagenode, Liège, Belgium) ultrasonication bath for 10 min at high energy setting. Lysates were clarified by centrifugation (20,000 \times g, 10 min) and protein concentration was quantified using the BCA assay kit (Pierce, Rockford, IL). In a typical experiment, 40 μg of total protein lysate was mixed with an equal amount of a $^{13}\text{C}_6$ - $^{15}\text{N}_2$ -lysine labeled lysate pool consisting of equal parts of lysates from WT animals aged 1, 6, 12 and 17 days. Proteins were reduced, alkylated and digested with endoproteinase LysC (Wako Bioproducts, Richmond, VA) using the FASP method (Wiśniewski et al., 2009). Peptide mixtures were either analyzed without fractionation or desalted via C18 solid phase extraction (SPE) cartridges (3M, St. Paul, MN) and subjected to isoelectric focusing on an OFFGEL system (Agilent, Santa Clara, CA) using 13 cm linear immobilized pH gradient strips with a pH range from 3 to 10 according to published procedures (de Godoy et al., 2008; Hubner et al., 2008). Fractionated or unfractionated peptides were purified via StageTips (Rappsilber et al., 2007).

Biochemical Isolation of Protein Aggregates

Approximately 600 worms were resuspended in 550 μl lysis buffer (50 mM Tris/HCl pH 8.0, 0.5 M NaCl, 4 mM EDTA, 1% (v/v) Igepal CA630, Complete protease inhibitor cocktail (Roche Diagnostics, Mannheim, Germany)) and sonicated for 8 min at 0°C in a Bioruptor sonication bath at high energy setting. Lysates were clarified (1,000 g, 1 min, 4°C) and protein content was adjusted to equal levels. For proteome measurements, a lysate pool of $^{13}\text{C}_6$ - $^{15}\text{N}_2$ -lysine labeled animals was mixed with each of the samples. Insoluble proteins were sedimented by ultracentrifugation (500,000 rcf, 4°C, 10 min) and subsequently washed twice with modified RIPA buffer (50 mM Tris/HCl pH 8.0, 0.15 M NaCl, 4 mM EDTA, 1% (v/v) Igepal CA630, 0.5% sodium deoxycholate, complete protease inhibitor cocktail) before solubilization in 2% SDS containing sample buffer for 10 min at 95°C. Proteins were resolved by SDS-PAGE and either analyzed by Coomassie staining or immunoblotting, or processed for MS analysis by in gel digestion and StageTip purification.

MS and Data Analysis

Peptides were separated on C18 reversed phase nano-HPLC columns (Nagaraj et al., 2011; Walther and Mann, 2011) with gradient durations of 140 or 280 min for fractionated or unfractionated samples, respectively, and sprayed online into LTQ-Orbitrap Velos or Orbitrap-Elite mass spectrometers (Thermo Fisher Scientific, Bremen, Germany) (Michalski et al., 2012; Olsen et al., 2009). In each scan cycle, fragmentation spectra of the 10 most intense peptide precursors in the survey scan were acquired in the higher-energy collisional dissociation (HCD) mode (Olsen et al., 2005). Raw data was processed in the MaxQuant software environment (Cox and Mann, 2008) and peak lists were searched with Andromeda (Cox and Mann, 2011) against a database containing the translation of all predicted proteins listed in Uniprot (release January 15, 2012) as well as a list of contaminants including commonly observed human keratins as well as the NCBI protein database of *E. coli* strain K12 (release date January 25, 2010). The minimal required peptide length was set to seven amino acids and both protein and peptide identifications were accepted at a false discovery rate of 1%.

Proteomics raw data and selected MaxQuant output files have been deposited to the ProteomeXchange Consortium (Vizcaino et al., 2014) via the PRIDE partner repository with the dataset identifier PXD001364.

Proteasome Activity Assays

Worms were lysed by ultrasonication in the presence of 2 mM ATP and proteasomal activity was performed as previously described using the fluorogenic substrate Z-Gly-Gly-Leu-AMC (Kisselev and Goldberg, 2005; Vilchez et al., 2012).

Immunostaining and Microscopy

WT or *daf-2* mutant worms expressing HSP-16.1::GFP were mounted on a 2% agar pad in 0.25 mM levamisole. A minimum of 20 worms were scored for each strain. To visualize actin, worms were fixed in acetone and stained with rhodamine-phalloidin (Molecular Probes R415). Fluorescence imaging was performed on a Zeiss Axiovert fluorescence microscope.

In Silico Aggregation Propensity Calculations

As the observed insoluble aggregates are unlikely to be all of amyloid nature, to predict the intrinsic aggregation propensity of proteins we adopted the recently developed CamSol method (Sormanni et al., 2015a), which avoids the bias of the Zyggregator method of predicting specific amyloid aggregation rates (Tartaglia and Vendruscolo, 2008). We started from the CamSol intrinsic solubility profile, which consists of a score for each residue along the sequence, and used it as a starting point to calculate a single aggregation propensity score ('Z-score') for each protein. The main principle defining this Z-score from the CamSol intrinsic profile is that the overall aggregation propensity of the protein should be proportional to the contribution given by the regions of the sequence that are aggregation-prone, attenuated by the regions that are aggregation-resistant, and normalized with respect to the protein length.

The CamSol intrinsic profile is calculated using a linear combination of physico-chemical properties of amino acids that have generally been associated with the solubility of proteins (Chiti et al., 2003, Fernandez-Escamilla et al., 2004, Pawar et al., 2005, Pechmann et al., 2009), namely hydrophobicity, electrostatic charge, and α -helical and β sheet propensities. The Wimley-White scale (Wimley and White, 1996) is selected as hydrophobicity scale and α -helix and β sheet propensities are calculated from the Protein Data Bank (PDB) using representative structures with a filter of 50% sequence identity (Sormanni et al., 2015a). The linear combination results in the position-dependent score s_i , which for a given residue i , is thus:

$$s_i = a_{hyd} p_i^{hyd} + a_c p_i^c + a_\alpha p_i^\alpha + a_\beta p_i^\beta \quad (1)$$

where p_i^{hyd} , p_i^c , p_i^α and p_i^β are the hydrophobicity, the charge (at neutral pH), the α -helical and the β sheet propensities, respectively, while the a values represent the constants of the linear combination (Sormanni et al., 2015a). The s_i values are then smoothed over a seven amino-acid window in order to account for the effects of neighboring residues, and corrected with two additional terms to provide the intrinsic solubility profile S_i

$$S_i = \frac{1}{7} \sum_{j=-3}^3 s_{i+j} + a_{pat} f_i^{pat} + a_{gk} f_i^{gk} \quad (2)$$

where f_i^{pat} is the correcting term that takes into account the presence of specific patterns of alternating hydrophobic and hydrophilic residues and f_i^{gk} is the correcting term that takes into account the gatekeeping effect of individual charges, defined as

$$f_i^{gk} = \sum_{j=-5}^5 e^{\frac{j}{200} c_{i+j}} \quad (3)$$

where c_{i+j} is the charge of the amino acid $i+j$.

Once the intrinsic CamSol profile S_i is obtained, the overall aggregation propensity score (Z-score) can be defined as

$$Z = - \frac{\sum_{i=1}^L \omega_i^{up} \tilde{S}_i^{up}(th^{up}) + \sum_{i=1}^L \omega_i^{down} \tilde{S}_i^{down}(th^{down})}{\gamma L^\delta}$$

where

$$\tilde{S}_i^{down}(th^{down}) = \begin{cases} S_i - th^{down} & \text{if } S_i < th^{down} \\ 0 & \text{if } S_i \geq th^{down} \end{cases}$$

is the residue-specific aggregation-prone contribution, which is dependent on the parameter th^{down} , that represents a minimum 'aggregating score threshold' filtering the regions of the profile that contribute to the overall aggregation propensity,

$$\tilde{S}_i^{up}(th^{up}) = \begin{cases} S_i - th^{up} & \text{if } S_i > th^{up} \\ 0 & \text{if } S_i \leq th^{up} \end{cases}$$

is the residue-specific aggregation-resistant contribution, which is dependent on th^{up} parameter, which is the threshold that filters the regions of the profile that act to attenuate the overall aggregation propensity, ω_i^{up} and ω_i^{down} are the soluble-region weights and aggregation-prone region weights respectively, L is the length of the protein and γ and δ are two constants that specify the functional law of the Z-score with respect to the protein length. The minus sign in the formula refers to our sign convention for which higher Z-scores represent higher aggregation propensity, while lower Z-scores indicate higher solubility. The parameters th^{up} , th^{down} , ω_i^{up} , ω_i^{down} , γ and δ , have been tuned with a Monte Carlo simulation aimed at maximizing both the correlation coefficient

of the Z-score with an ensemble of protein aggregation rates measurements available in the literature (DuBay et al., 2004), and the efficiency in the separation of the Z-score distributions of two datasets of known non-aggregating and aggregating peptides respectively, obtained from a systematic literature search.

Accuracy in the binary prediction of aggregate and non-aggregate peptide datasets has been tested upon finishing the optimization, and an additional correlation control has been performed with the solubility of a set of protein mutants (Sormanni et al., 2015a).

Statistics and Bioinformatic Analysis

Prediction of subcellular localization, signal sequences and transmembrane segments were performed using WoLF PSORT (Horton et al., 2007), SignalP (Petersen et al., 2011) and TMHMM v. 2.0 (Krogh et al., 2001) algorithms, respectively. Further annotation included predicted tissue specificity of expression (Chikina et al., 2009), analysis of Pfam protein families (Finn et al., 2008) and gene ontology databases (Ashburner et al., 2000). Benjamini-Hochberg FDR-controlled Fisher Exact test as well as one- and two-dimensional annotation enrichment analysis was performed in the Perseus data analysis suite (Cox and Mann, 2012). Fuzzy c-means clustering of time course profiles was carried out using the Mfuzz package in the statistical programming language R (Kumar and Futschik, 2007).

In time course analyses of individual proteins in WT animals, only those proteins were displayed that were quantified at day 1 and at least at three consecutive time points. Statistical significance of abundance differences of protein subsets across different time points or strains were generally performed using the Wilcoxon signed rank test in which only those proteins were considered that were quantified in both conditions. To identify aggregation-prone proteins that were significantly affected by aging, those proteins that were quantified in at least 3 out of 4 biological replicate experiments at day 1 and day 12 were subjected to a Welch's t test and filtered based on a 5% permutation-based false discovery rate threshold.

SUPPLEMENTAL REFERENCES

- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al.; The Gene Ontology Consortium (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29.
- Brenner, S. (1974). The genetics of *Caenorhabditis elegans*. *Genetics* 77, 71–94.
- Chiti, F., Stefani, M., Taddei, N., Ramponi, G., and Dobson, C.M. (2003). Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* 424, 805–808.
- Cox, J., and Mann, M. (2012). 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. *BMC Bioinformatics* 13 (Suppl 16), S12.
- de Godoy, L.M.F., Olsen, J.V., Cox, J., Nielsen, M.L., Hubner, N.C., Fröhlich, F., Walther, T.C., and Mann, M. (2008). Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* 455, 1251–1254.
- DuBay, K.F., Pawar, A.P., Chiti, F., Zurdo, J., Dobson, C.M., and Vendruscolo, M. (2004). Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains. *J. Mol. Biol.* 341, 1317–1326.
- Fernandez-Escamilla, A.-M., Rousseau, F., Schymkowitz, J., and Serrano, L. (2004). Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.* 22, 1302–1306.
- Finn, R.D., Tate, J., Mistry, J., Coghill, P.C., Sammut, S.J., Hotz, H.R., Ceric, G., Forslund, K., Eddy, S.R., Sonnhammer, E.L., and Bateman, A. (2008). The Pfam protein families database. *Nucleic Acids Res.* 36, D281–D288.
- Haslbeck, V., Eckl, J.M., Kaiser, C.J., Papsdorf, K., Hessling, M., and Richter, K. (2013). Chaperone-interacting TPR proteins in *Caenorhabditis elegans*. *J. Mol. Biol.* 425, 2922–2939.
- Horton, P., Park, K.J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., and Nakai, K. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* 35, W585–W587.
- Hubner, N.C., Ren, S., and Mann, M. (2008). Peptide separation with immobilized pi strips is an attractive alternative to in-gel protein digestion for proteome analysis. *Proteomics* 8, 4862–4872.
- Kisselev, A.F., and Goldberg, A.L. (2005). Monitoring activity and inhibition of 26S proteasomes with fluorogenic peptide substrates. *Methods Enzymol.* 398, 364–378.
- Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580.
- Michalski, A., Damoc, E., Lange, O., Denisov, E., Nolting, D., Muller, M., Viner, R., Schwartz, J., Remes, P., Belford, M., et al. (2012). Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes. *Mol. Cell. Proteomics* 11, O111.013698.
- Nagaraj, N., Wisniewski, J.R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Pääbo, S., and Mann, M. (2011). Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* 7, 548.
- Olsen, J.V., de Godoy, L.M., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S., and Mann, M. (2005). Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* 4, 2010–2021.
- Olsen, J.V., Nielsen, M.L., Damoc, N.E., Griep-Raming, J., Moehring, T., Makarov, A., Schwartz, J., Horning, S., and Mann, M. (2009). Characterization of the Velos, an Enhanced LTQ Orbitrap, for Proteomics. *Mol. Cell. Proteomics* 8, S40.
- Pawar, A.P., DuBay, K.F., Zurdo, J., Chiti, F., Vendruscolo, M., and Dobson, C.M. (2005). Prediction of “aggregation-prone” and “aggregation-susceptible” regions in proteins associated with neurodegenerative diseases. *J. Mol. Biol.* 350, 379–392.

The logo for the journal Cell, consisting of a blue square with the word "Cell" in white text.

Pechmann, S., Levy, E.D., Tartaglia, G.G., and Vendruscolo, M. (2009). Physicochemical principles that regulate the competition between functional and dysfunctional association of proteins. *Proc. Natl. Acad. Sci. USA* 106, 10159–10164.

Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786.

Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* 2, 1896–1906.

Tartaglia, G.G., and Vendruscolo, M. (2008). The Zyggregator method for predicting protein aggregation propensities. *Chem. Soc. Rev.* 37, 1395–1401.

Vizcaino, J.A., Deutsch, E.W., Wang, R., Csordas, A., Reisinger, F., Ríos, D., Dianas, J.A., Sun, Z., Farrah, T., Bandeira, N., et al. (2014). ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* 32, 223–226.

Wimley, W.C., and White, S.H. (1996). Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Biol.* 3, 842–848.

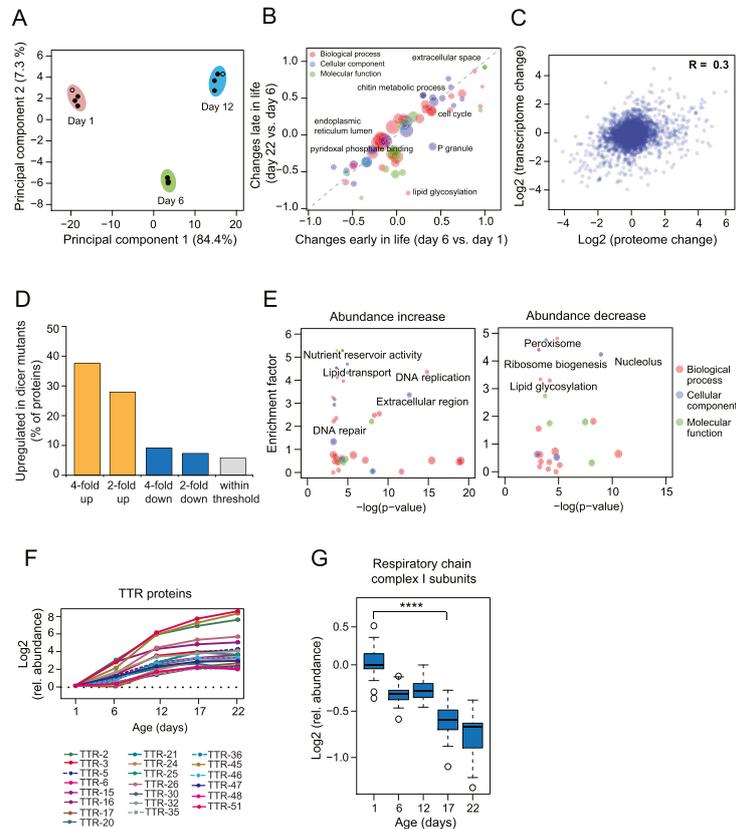


Figure S1. Proteomic Analysis of Aging in *C. elegans*, Related to Figure 1

(A) Reproducibility of SILAC-based proteomic analyses. Four biological replicates of WT animals at the indicated ages were collected independently and their proteomes were analyzed by quantification against the same SILAC spike-in standard. Each one of the four replicates (open circles) was prepared on a different day than the remaining three (closed circles). Principal component analysis shows that different individuals of the same age cluster closely together in the two-dimensional plot, demonstrating that the sum of technical and biological variation is much smaller than the age-related changes in the *C. elegans* proteome (Table S1A). On a whole proteome level, this allows analyzing results as a function of age only, and to disregard technical variation and differences between individual worms.

(B) Comparison of GO categories of proteins that changed in abundance in the early (day 6 versus day 1) and late stage in life (day 22 versus day 6) of WT animals (Table S2A). All terms that were significantly affected in either of the two periods are displayed (Wilcoxon rank sum test at 5% Benjamini Hochber FDR) and their relative changes were plotted against each other. The dashed gray line indicates the position of categories which are equally affected early and late in life. Selected outliers are indicated in the plot.

(C) Correlation between transcriptome data (Golden and Melov, 2004) and the proteomics dataset of this study. The Pearson correlation R between both datasets is displayed.

(D) Proteins that increase in abundance during aging are targets for dicer-mediated miRNA repression (Welker et al., 2007). The fractions of dicer (*dcr-1*) targets (transcriptionally upregulated in *dcr-1* mutants) among proteins that increased > 4-fold (50 of 133 proteins) or > 2-fold (99 of 357 proteins) in abundance from day 6 to day 22 of aging as well as proteins that decreased < 4-fold (6 of 66 proteins) or < 2-fold (25 of 325 proteins) in abundance or remained within threshold (less than 2-fold change in either direction) (173 of 3307 protein) are shown.

(E) Significantly affected functional categories among the proteins that increased 2-fold (left panel) or decreased 2-fold (right panel) in abundance in aged (day 22) animals. The enrichment factors of gene ontology (GO) terms are plotted against the p-value derived from Fisher Exact tests. Each term is represented by a circle. The size of the circle reflects the number of proteins affected. Only categories with at least 4 members are displayed. Selected categories are indicated (Tables S2B and S2C).

(F) Abundance change of TTR-like proteins during aging.

(G) Abundance change of quantified subunits of the mitochondrial respiratory chain complex I during aging. At least 17 subunits were quantified at each time point. ****p-value < 1.53×10^{-15} from Wilcoxon signed rank test.

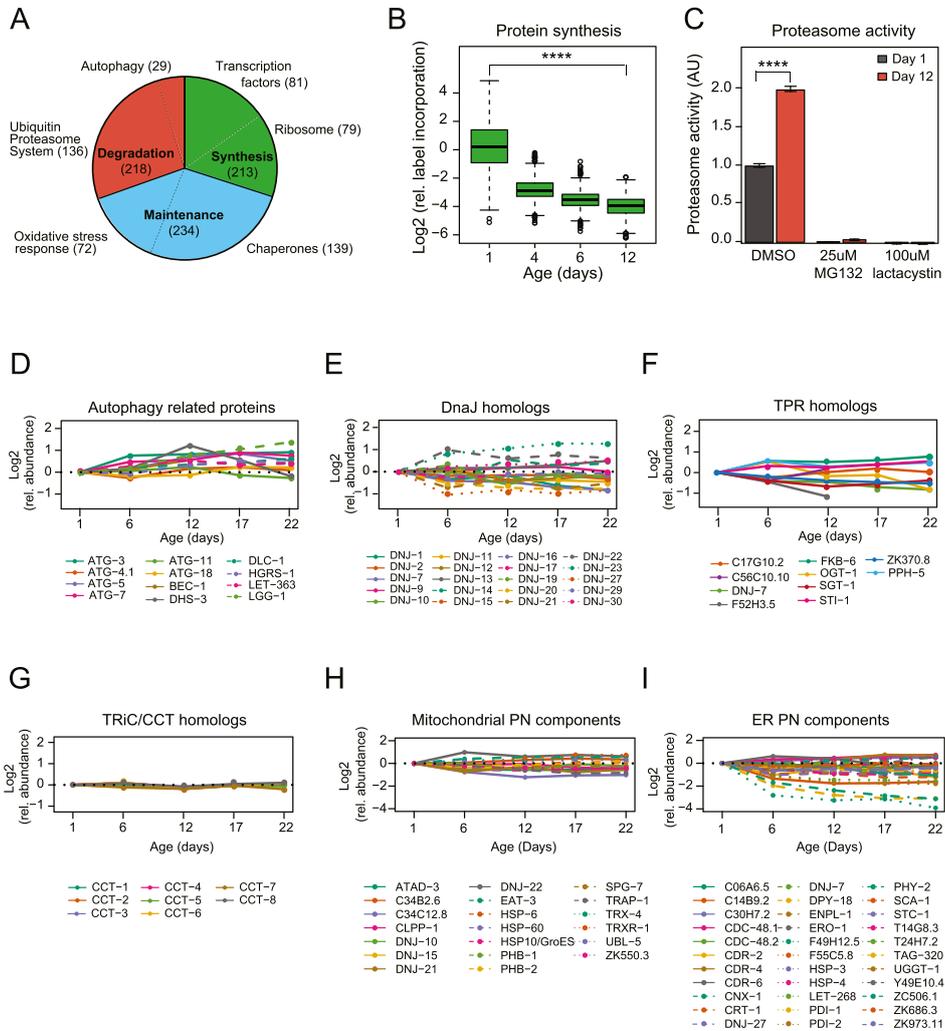


Figure S2. Abundance Changes in Components of the Proteostasis Network during Aging in WT Animals, Related to Figure 2

(A) The proteostasis network (PN) is divided into three main categories: protein synthesis (green), comprising transcription factors, translation factors and ribosomal components; conformational maintenance (blue), comprising components involved in folding and stress response; and degradation (red), comprising components involved in the ubiquitin proteasome system and autophagy (Table S3A).

(B) Decline of protein synthesis during aging determined by SILAC pulse labeling. Animals aged 1, 4, 6 or 12 days were transferred to a heavy lysine-labeled food source for 24 hr. Boxplot shows the distribution of heavy lysine incorporation into proteins. At least 500 proteins were quantified at each time point (Table S1C). p-value $< 2.2 \times 10^{-16}$ from Wilcoxon signed rank test.

(C) Chymotryptic proteasome activity in lysates of old (day 12) and young (day 1) worms, as measured with fluorogenic synthetic peptide as substrate in the presence of ATP (see Extended Experimental Procedures). Assays were performed in the absence (DMSO) or presence of proteasome inhibitors lactacystin (100 μ M) or MG132 (25 μ M). Error bars represent standard deviations from 6 replicate experiments. p-value $< 7.4 \times 10^{-13}$ from Welch's t test.

(D-I) Abundance change of proteins involved in protein folding and stress response in WT. Autophagy-related components (D), DnaJ/Hsp40 homologs (E), TPR domain proteins that potentially interact with chaperones (Haslbeck et al., 2013) (F), subunits of the TRiC/CCT chaperonin (G), components of the mitochondrial PN (H), components of the endoplasmic reticulum PN (I) (Table S3B).

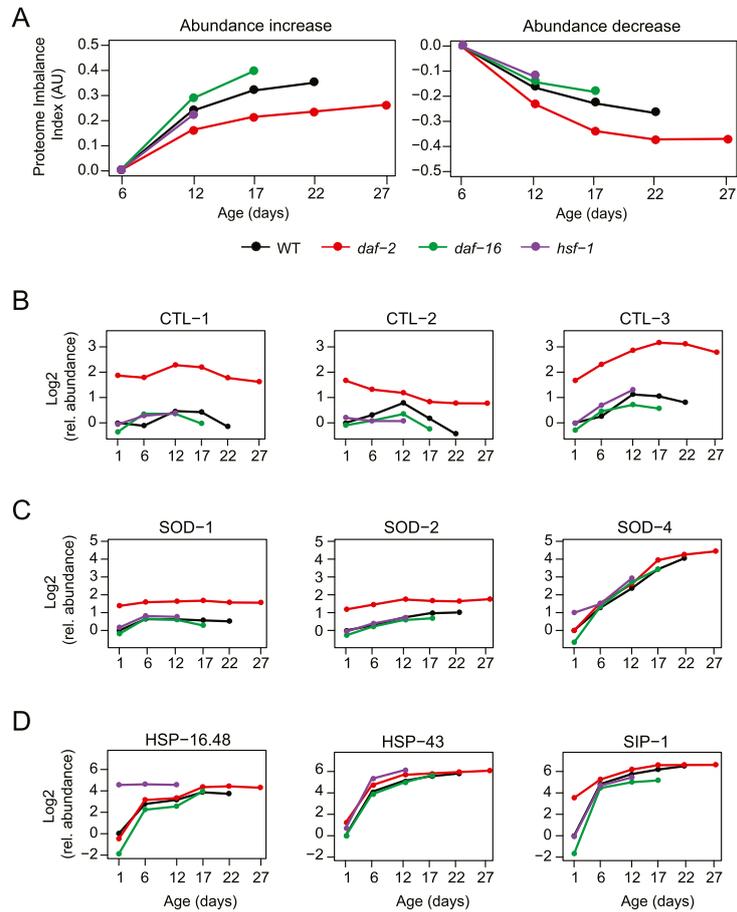


Figure S3. Proteome Changes in WT and IIS Mutant Worms, Related to Figure 3

(A) Proteome imbalance in WT, *daf-2*, *daf-16* and *hsf-1* mutant animals expressed as proteome imbalance index. Abundance differences of proteins that increased (left) or decreased (right) in abundance during aging relative to day 6 were summed up for each strain and normalized to the number of quantified proteins. The total number of quantified proteins was similar in the different worm strains and ranged from ~3743 to ~4700 proteins (Table S1B).

(B–E) Abundance profiles of catalases (B), SOD proteins (C), and small HSPs (D) along the lifespan of WT, *daf-2*, *daf-16* and *hsf-1* animals. Log₂ changes in abundance are shown relative to WT animals at day 1.

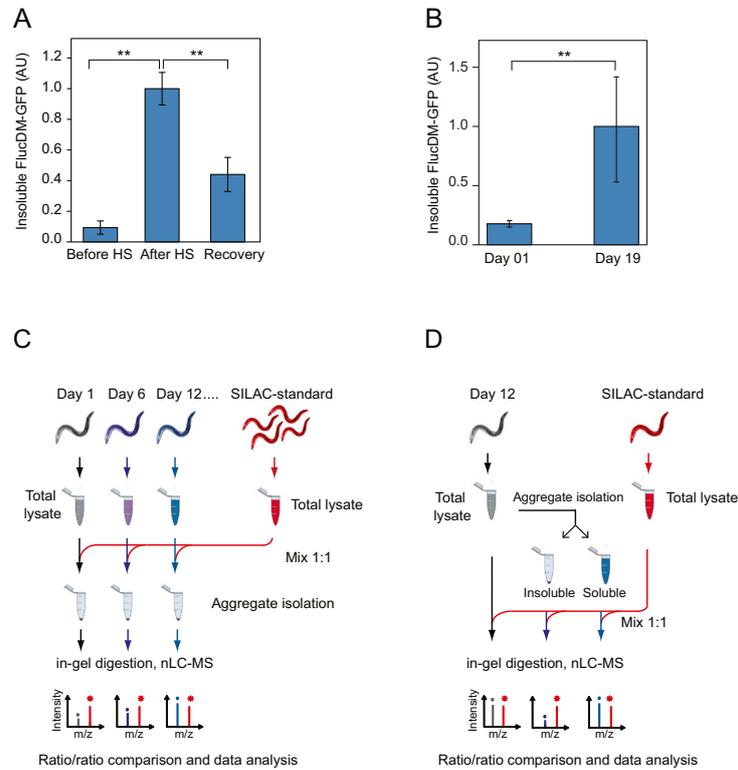


Figure S4. Isolation of Insoluble Aggregates from Worm Lysates, Related to Figure 4

(A and B) Validation of the biochemical procedure for the isolation of insoluble protein aggregates was performed with WT worms expressing FlucDM-GFP, a destabilized double mutant of firefly luciferase fused to GFP (Gupta et al., 2011) in body wall muscle cells.

(A) FlucDM-GFP expressing worms (day 1) were exposed to heat shock (33°C for 90 min), followed by recovery for 90 min at 20°C. Insoluble fractions were isolated from worm lysates by centrifugation (see Experimental Procedures), analyzed by immunoblotting with anti-luciferase antibodies, and quantified by densitometry. Error bars represent standard deviations from three independent experiments. **p-value < 0.01, Welch's t test.

(B) Fraction of insoluble FlucDM-GFP in young (day 1) and old (day 19) WT worm populations analyzed as above. Error bars represent standard deviations from four independent experiments. **p-value < 0.01, Welch's t test.

(C) Experimental design for the quantitative analysis of insoluble proteins. Synchronized worm populations at different ages were lysed and mixed with a metabolically (SILAC) labeled internal protein standard. Insoluble proteins were isolated, separated by SDS-PAGE, subjected to in gel digestion and analyzed by nano-HPLC coupled MS.

(D) Experimental design for the analysis of protein aggregation propensities. Lysates from aged worms (day 12) were fractionated. Total lysates as well as insoluble and soluble fractions were quantified against an identical SILAC standard to determine the insoluble and soluble fraction of each protein.

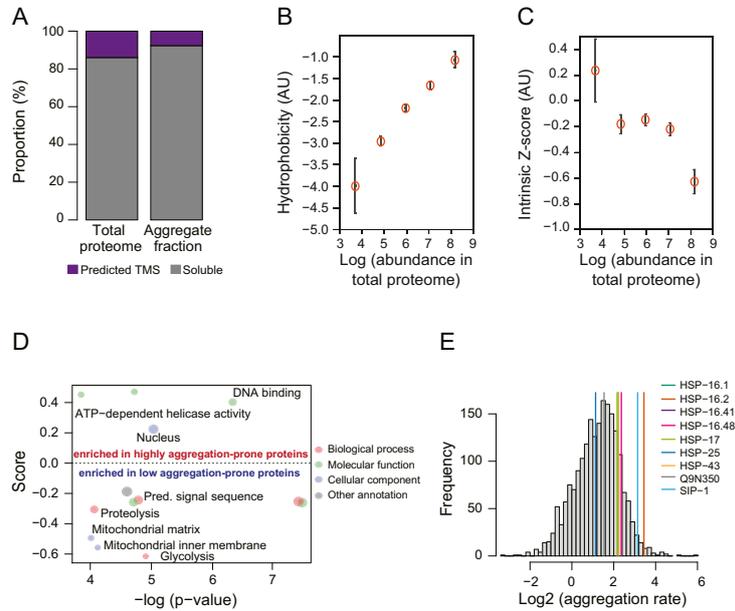


Figure S5. Proteomic Analysis of Protein Aggregation during Aging, Related to Figure 4

(A) Procedure for aggregate isolation does not enrich for membrane proteins. Proportions of identified proteins in total and insoluble fractions for soluble proteins and proteins that were predicted to contain at least one transmembrane segment (TMS) are shown. TMS were predicted with THMM.

(B and C) Physico-chemical properties of proteins that affect aggregation behavior. The overall hydrophobicity (B) of proteins (which is associated with a tendency of forming a stabilizing hydrophobic core) increases with their abundance, and the aggregation propensity score (Z-score; see [Extended Experimental Procedures](#)) of the proteins (C) decreases with their abundance, indicating that highly abundant proteins tend to be more soluble.

(D) GO annotation distribution among proteins with high and low intrinsic aggregation propensities in WT animals at day 12. Enrichment scores and p-values of a Wilcoxon rank sum test are plotted against each other. Only categories with a cutoff of 5% Benjamini-Hochberg FDR and at least 4 proteins are displayed ([Table S4A](#)).

(E) Members of the small HSP family of molecular chaperones display a high aggregation rate compared to the overall distribution of aggregation propensities of all quantified proteins in the proteome of day 12 WT animals (Wilcoxon rank sum test p-value: 3.4×10^{-3}).

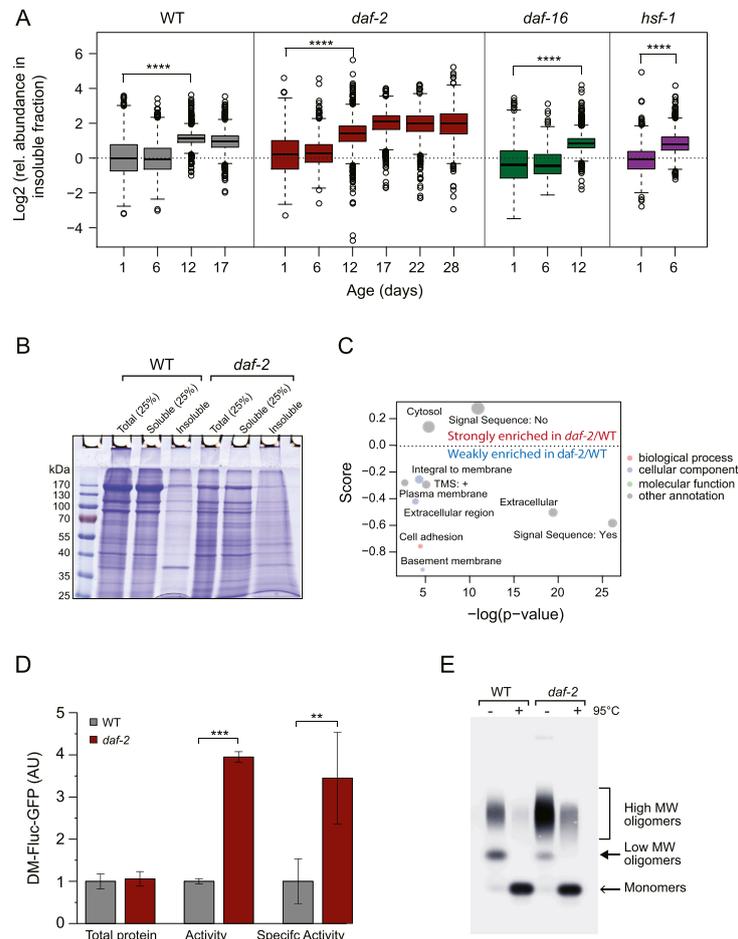


Figure S6. Protein Aggregation in Lifespan Mutant Worms during Aging, Related to Figure 5

(A) Time course analysis of protein aggregation in *daf-2*, *daf-16* and *hsf-1* mutants, compared to WT. Boxplots of SILAC ratios as in Figure 5A are shown. In this experiment fractionation of *hsf-1* mutant animals of 12 days of age was not reliable due to the limited number of live worms recovered. Number of quantified proteins: WT day 1, 2010; day 6, 1698; day 12, 1987; day 17, 1355; *daf-2* mutant day 1, 1093; day 6, 1103; day 12, 2660; day 17, 1599; day 22, 1572; day 28, 1759; *daf-16* mutant day 1, 1120; day 6, 993; day 12, 1366; *hsf-1* mutant day 1, 1318; day 6, 1739 (Table S1D). ****p-value < 2.2×10^{-16} from Wilcoxon signed rank test.

(B) Analysis of differential aggregation between aged *daf-2* mutants and WT animals (day 12). Worm samples were separated into soluble and insoluble fractions and analyzed by SDS-PAGE and Coomassie staining. Note that only 25% of total and soluble fraction is loaded. One representative out of 3 independent experiments is shown.

(C) Annotation distribution analysis of differential aggregation between *daf-2* and WT animals at 12 days of age. Enrichment scores and p-values of a Wilcoxon rank sum test are plotted against each other (Table S4B).

(D) Increased proteostasis capacity in *daf-2* mutant worms. Fluc-DM-GFP and luciferase activities were measured by immunoblotting with anti-luciferase antibody and luminescence assay, respectively, in 12 days old WT and *daf-2* mutant worms expressing muscle-specific Fluc-DM-GFP. Protein levels, activities and specific activities (activities per amount of FlucDM-GFP protein) are given in arbitrary units with values in WT set to 1. Results from 7 independent measurements \pm SD are shown. ***p-value < 0.0001; **p-value < 0.001, Welch's t test.

(E) SDD-AGE analysis of WT and *daf-2* mutant worms expressing muscle-specific Q35-GFP. Worm extracts from 2 days old animals were analyzed with and without incubation at 95°C for 10 min. Q35-GFP was visualized by immunoblotting. Note that total amounts of Q35-GFP detected after heat treatment are similar in WT and *daf-2* mutant extracts.

2.4 Importance of the analysis

Experimental data of MS-based protein abundances shows that widespread protein aggregation occurs upon ageing in WT nematodes as the proteostasis machinery gets impaired and the proteome remodelled (Figure 4 and 1 in section 2.3). The phenomenon of protein aggregation was already largely associated with pathological events, namely misfolding and neurodegenerative diseases [59, 62, 67, 96, 230]. Intuitively, since we observe in WT animals that the amount of protein aggregates increases upon ageing (Figure 4A, section 2.3), we would expect the long-lived strain *daf-2* to show a lower amount of inclusions than the WT when comparing age-matched data points. Conversely, we would logically expect the short-lived mutant strain, *hsf-1*, to accumulate more aggregates than the WT upon ageing. However, both long-lived and short-lived mutants exhibit the presence of more insoluble protein deposits than the WT strain as the animals grow older instead (Figure 5A and S6A, section 2.3). Aggregation is also widespread in long-lived and short-lived strains, with more than a thousand proteins forming insoluble aggregates. This translated in a large overlap between the proteins identified as forming deposits, however the extent by which each protein would aggregate was very diverse across the strains (Figure 5 and S6, section 2.3).

To understand the differences in the aggregation observed in WT, short-lived and long-lived strains, we asked whether the overall extent of insoluble material accumulated differently across the strains could be explained in terms of some specific, general physico-chemical principles. Since both short-lived and long-lived animals were forming more aggregates than WT at day 12 of adulthood (when proteostasis decline is observed in the WT strain), we separated in the analysis the proteins forming less aggregates in *hsf-1* (short-lived) compared to WT and *daf-2* (long-lived) compared to WT. For all the remaining proteins in *daf-2* or *hsf-1* forming more aggregates than WT at day 12, we looked for a collective differential behaviour in terms of aggregated mass reflected in a given physico-chemical property. For each mutant strain, long- or short-lived, we grouped the proteins forming more aggregates

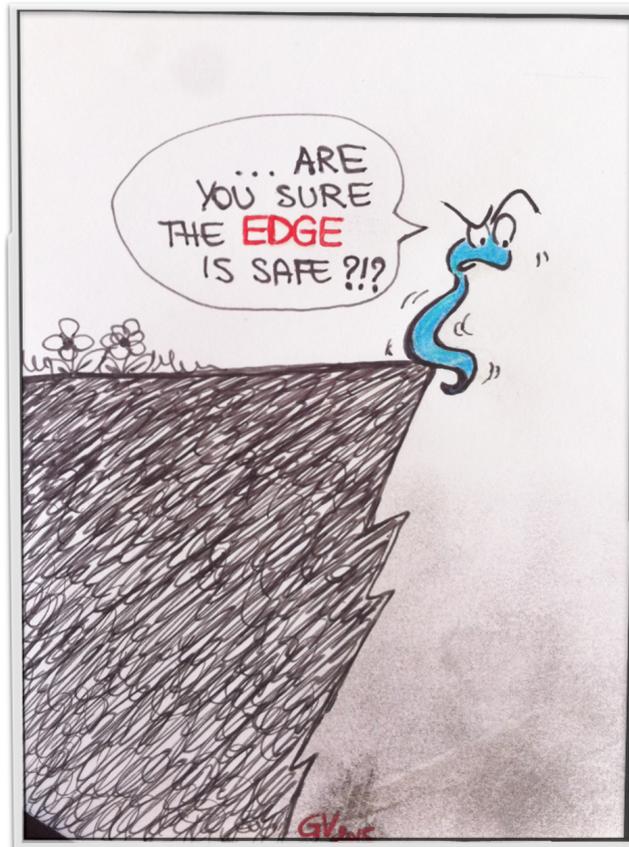
into four quartile groups, according to the fold difference in their aggregated mass with respect to the WT strain. In each group, we then evaluated aggregation-related physico-chemical properties using the Camsol Intrinsic method discussed in section 2.2 and 2.3 and the s2D predictor of secondary structure populations [209]. This enabled us to highlight which kind of properties may drive the accumulation of insoluble aggregates in the long-lived and short-lived strains compared to the WT animals.

In Figure 5 D-G of section 2.3 we report the physico-chemical properties that we found were significantly modulated by the amount of insoluble aggregates present in either the long- or short-lived strains compared to the WT. Strikingly, a completely different pattern emerges when we compare the aggregation increase in the *daf-2* strain and in the *hsf-1* strain against the WT in terms of physico-chemical principles. While proteins aggregating the most in the short-lived strain compared to the WT are the most aggregation prone (Figure 5D, right scatter plot with violet points), this pattern is reversed in the long-lived strain (Figure 5D, left scatter plot with red points). Also, proteins forming increasingly more aggregates in the *daf-2* strain compared to the WT are more charged, disordered and hydrophilic (Figure 5E-G, section 2.3), suggesting that the enhanced aggregation observed in this long-lived strain is not the result of an intrinsic pathological process accompanying proteostasis decline. Instead, it could be the result of an extrinsic mechanism triggered by the organism to remove soluble, toxic oligomeric species by assembling them into larger insoluble deposits, in order to restore protein homeostasis.

Combining the experimental information of the protein aggregate abundances with the *in silico* physico-chemical properties predictions thus represented a key step in understanding the molecular bases of proteostasis disruption and its restoring. Furthermore, it provided the first pieces of evidence, in a living organism, of a protective proteome-wide aggregate response activated upon ageing to restore functional protein balance.

Chapter 3

*Life Over the Solubility Edge in ageing *C. elegans**



3.1 Introduction

Almost a decade ago, Tartaglia and co-workers proposed the existence of a strong anticorrelation between the aggregation rates, which were measures *in vitro* of the limited group of proteins available in the literature at the time, and the corresponding human mRNA expression levels measured *in vivo* [130]. This anticorrelation was explained as a net result of two opposing effects on the primary sequence of a protein. On one side, the effect of random mutations, which tend to increase the aggregation propensity of the protein; and on the other side the effect of evolution, which push to keep the protein soluble at the concentration required in the cell for their biological roles [231–235]. Hence, proteins would have co-evolved with their cellular environment to be sufficiently soluble at the levels at which they need to be expressed in the cell for optimal functioning, therefore not much affected by the cellular macromolecular crowding [236], but with almost no margin of safety to respond to genetic or environmental factors that either decrease their solubility or increase their cellular concentration [130]. This co-evolution involves the presence of a complex network of cellular machineries, the protein homeostasis network, responsible for maintaining proteome functional balance against aberrant folding and aggregation, processes that can generate toxic species in the cell [72, 74, 77, 237]. As organisms grow old, however, the protein homeostasis network becomes impaired and widespread aggregation occurs [90, 94, 95, 238]. Protein aggregation in turns further disrupt protein homeostasis [68], forming a vicious circle of degeneration [65, 67, 72, 74, 77, 82]. As already seen in Chapter 1, age is considered to be the greatest risk factor for a variety of neurocognitive disorders, which represent the most debilitating, expensive, and common medical conditions in recent times [24, 34] and whose nearly universal hallmark is the formation of protein aggregates [61–63].

The concept by which proteins are expressed close to their solubility limits has been referred to as the "life on the edge hypothesis" [130]. Suggested on the basis of the limited experimental data available at the time, whether this concept could have general validity for proteins in the cell remains an open question, which could

help rationalize a number of observations on protein aggregation in ageing and neurodegenerative diseases.

Using mass spectrometry data [90], here we provide the first proteome-wide evidence for the life on the edge hypothesis in the nematode *C. elegans*, and we show that proteins in the adult worm are actually living just over the edge of their solubility. To determine to what extent proteins cross the solubility limit, we classify three measured physical quantities related with protein levels, namely abundance, aggregation and supersaturation, and we link them to the life on the edge hypothesis. The observation that proteins are found in the adult nematode just over the solubility limit prompts the question of how this limit is crossed upon ageing and if it is a general threshold over which any protein would aggregate. To tackle these questions, we show that with age there is a proliferation in the aggregate levels, which is not due to an increase in the overall protein content in the worm, suggesting that the critical concentration is a threshold that is protein-specific and that aggregation upon age may be due to a decrease in the solubility rather than an increase in the expression levels. By looking at the different contributions that supersaturation, abundance and aggregation give to this age-dependent proliferation in aggregate levels, we suggest that supersaturation plays a key role, specifically for the most abundant proteins, in the process of healthy ageing, as the largest contribution to the total aggregate mass is provided by the most abundant proteins. Further studies will be needed to establish how these principles can be extended from healthy ageing to neurodegenerative disorders, and from worms to humans.

3.2 Results

3.2.1 Life on and over the solubility edge in adult *C. elegans*

In order to test whether the life on the edge hypothesis holds true at a proteomic level in an eukaryotic organism, we took advantage of recent in vivo proteome-wide mass spectrometry data acquired in wild-type strain nematodes *C. elegans* focusing

on one set up experiment in which total, soluble (supernatant) and insoluble (pellet) protein abundances were measured in the adult worm [90]. By looking at the proteins detected in at least two out of three replicas, we found that in the adult wild-type worm (day 12 of adulthood from L4 stage), about 75% of the proteins detected are found in the pellet fraction (Figure 3.1A, 2792 proteins). This observation implies that nearly three quarters of all proteins detected is expressed at least at the critical concentration. Of these proteins, we evaluated and normalised the total, supernatant and pellet abundance in terms of mass spectrometry absolute LFQ values (see section 3.4.1). Defining the solubility as the supernatant abundance in the presence of a pellet fraction, we observe that in adult worms nearly all proteins are lying on, or in fact just above, the edge of their solubility (Figure 3.1B).

The findings that we have reported here, therefore, suggest that these data represent experimental evidence for the life on the edge hypothesis in an eukaryotic model organism and indicate that widespread aggregation occurs, although the amounts of aggregates are small in comparison to the total abundance. Such finding is also in agreement with the trend previously observed for the *E. coli* proteome in a cell-free reconstituted system [239].

3.2.2 Life over the solubility edge: the relationship between aggregation, supersaturation and abundance of proteins

Since nearly all detected proteins in the adult worm are expressed on and just above their solubility limits (Figure 3.1) widespread aggregation occurs, thus compromising protein homeostasis. We next asked how can we measure, from the data, the extent and the way in which a protein is exceeding the solubility limit. To tackle this question, we define three main ways in which the landscape over the solubility edge, represented by the Pellet-Soluble abundance plane (Figure 3.2) can be spanned and analysed in terms of risk to the cellular protein homeostasis. Each way is characterised by the evaluation of a physical quantity whose value increases in a different direction with respect to the solubility limit.

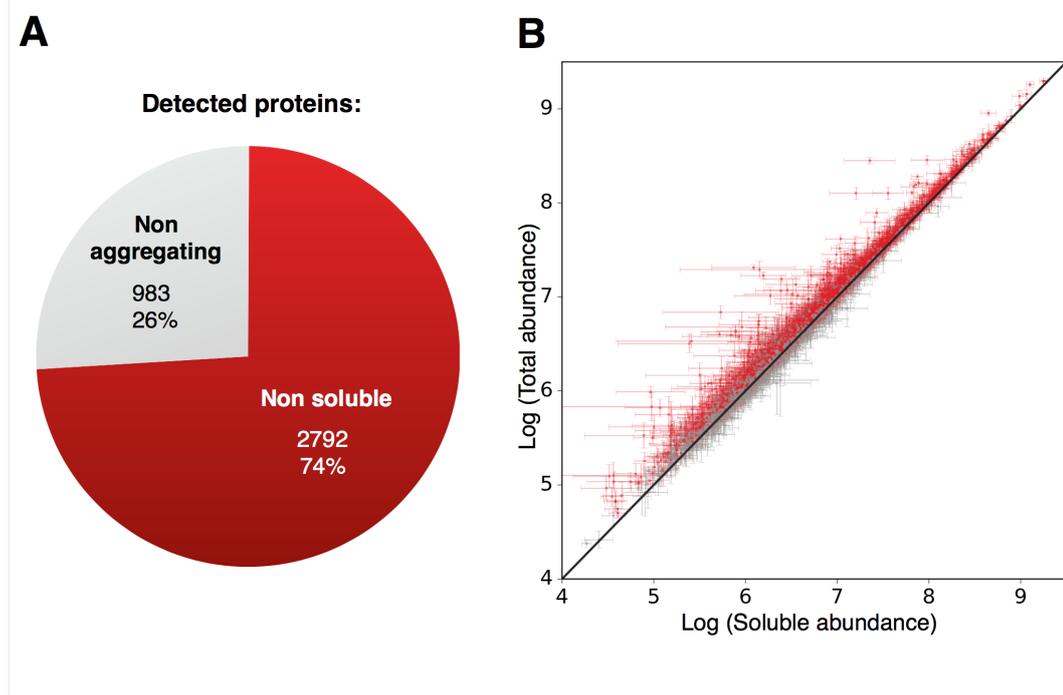


Fig. 3.1
Validation of the Life on the Edge hypothesis in adult worms

(A) Pie chart showing the percentage and number of aggregating (red filling) and non-aggregating proteins detected in aged (day 12) wild-type worms. (B) Scatterplot of protein total abundance (T) plotted against soluble abundance (S) in \log_{10} scale for each protein quantified (see section 3.4.1). The colour code is matching the pie chart choice. The unity line, corresponding to the solubility limit for all proteins detected with an aggregated fraction (red points), is shown in black. For the proteins that do not form insoluble deposits (grey points), we do not have an estimate of the solubility limit, but only a lower bound for it (i.e. the total abundance).

The first quantity is supersaturation (σ), which is defined as the ratio between the total abundance (T) of a protein and its critical concentration C^* , which can be approximated as the supernatant abundance (S) in the presence of a non-negligible pellet fraction (P):

$$\sigma = \frac{T}{C^*} = \frac{T}{S} \Big|_{P \neq 0} \quad (3.1)$$

With such a definition, supersaturation represents an experimental index of aggregation propensity for a protein. The more the protein is supersaturated, the more it is prone to aggregate.

The second quantity is the protein aggregate load, defined as the pellet abundance P (see section 3.4.1) and satisfying

$$P = T - S \quad (3.2)$$

This variable gives information on the total mass of the aggregates of a given protein. The third observable is the total protein abundance in presence of a detectable aggregate load, $(T|_{P \neq 0})$, which is a quantity proportional to the cellular protein expression.

We evaluated supersaturation σ , aggregate abundance P and total abundance T for all the 1828 proteins (see section 3.4.1) reported in the life on the edge plot (Figure 3.1B, red points). These values refer to the proteome situation in the adult day-12 nematode.

Since each quantity is maximised in a different direction with respect to the solubility edge, different regions can be identified in the life over the solubility edge landscape where proteins populate high values of supersaturation, aggregate or total abundance. In order to show these regions and their overlaps, we highlighted the 20% most saturated, 20% most aggregated and 20% most abundant proteins (365 proteins in each set) with different colours in Figure 3.2 and used a subtractive colour mixing model for their overlap, reporting a Venn diagram with the number of proteins present in each intersection. Supersaturation, aggregation and abundance have blue, yellow and red as base colours respectively.

We identified six different classes from the overlaps of the most supersaturated, aggregated and abundant groups. We found that 290 proteins are only most supersaturated (blue points in Figure 3.2), 34 proteins are only most aggregated (yellow points in Figure 3.2) and 81 proteins are only most abundant (red points in Figure 3.2). Most abundant and most aggregated but not highly supersaturated proteins are represented in orange in Figure 3.2 (256 entries), while 47 proteins appears to be most supersaturated and most aggregated but not highly abundant (green points in

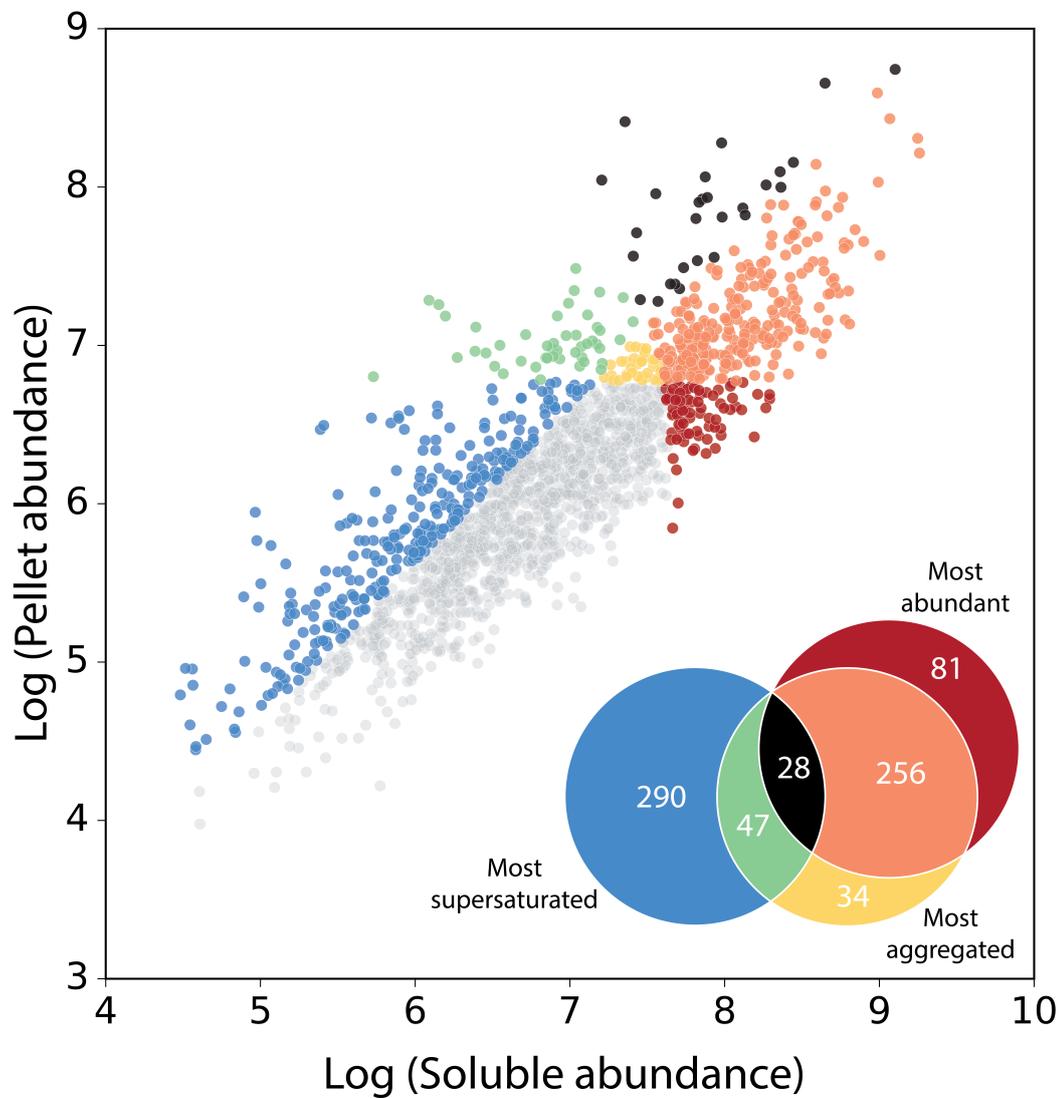


Fig. 3.2

Visualisation of the most supersaturated, aggregated and abundant proteins in the life over the edge solubility landscape

Scatterplot of pellet abundance (P) versus soluble abundance (S) in \log_{10} scale (see section 3.4.1) showing in colours the three classes of top 20% most supersaturated, most abundant and most aggregated proteins and their intersections. Each class contains 365 proteins. Number of proteins in each subgroup and relative colour code are reported in the Venn diagram. Abundance data refer to day 12 wild-type worms.

Figure 3.2). The black points in Figure 3.2, corresponding to 28 proteins, belong to the region in which all the quantities have maximal values. Of the 1828 proteins shown in the scatterplot of Figure 3.2, 1092 do not belong to any of the three sets, and are shown in grey.

The overlaps between the most abundant, supersaturated and aggregated sets provide information about the relationship between our physical quantities. While there is a high overlap between the most aggregated and the most abundant proteins, the most supersaturated set does not share a high number of proteins. Since supersaturation represent an index of the aggregation propensity, we suggest that a vast majority of the high abundant proteins have evolved to be less aggregation prone, which is in agreement to what we predicted with our sequence-based algorithm [90]. This trend is in agreement with previous findings that report an evolutionary conserved trend of aggregation-prone proteins to be low abundant and subject to decrease synthesis and high turnover [240]. However, the degree by which they are more soluble is not high enough to avoid them from aggregating more than the low abundant proteins.

On the other hand, the black proteins of Figure 3.2 (28 points) are highly abundant and highly aggregation prone, therefore they also end up aggregating the most. Such proteins, thus, represent a metastable subproteome in which the evolutionary pressure might not be applied in the adult worm.

3.2.3 Aggregate proliferation in ageing *C. elegans*

Having established that proteins can cross over the solubility edge in terms of supersaturation, abundance or aggregation, we now wish to probe if the critical concentration is a property that is protein dependent, or if there is an overall critical concentration over which all proteins aggregate. To address this question, we used SILAC and absolute LFQ measurements of mass spectrometry total and pellet protein abundance data upon ageing [90] to evaluate both the total and insoluble mass variation of the proteome in the ageing *C. elegans* (Figure 3.3). Proteins predicted to

be extracellular (see section 3.4.2) were excluded from the calculations, since being functional in the extracellular space should have induced less evolutionary pressure on them to evolve to higher solubilities relative to the intracellular proteins.

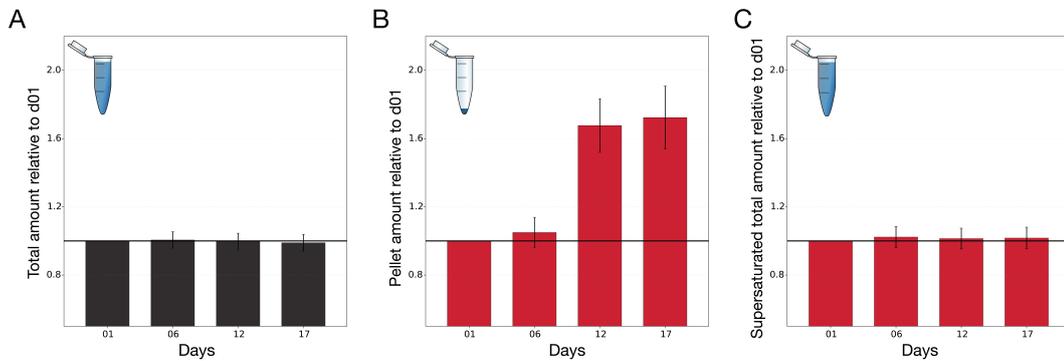


Fig. 3.3
Evolution of total and pellet amounts in ageing worms

Bar charts of the variations in intracellular proteome amounts upon ageing in wild-type worms. The sum of the contributions of intracellular proteins relative abundances are estimated at various time points with respect to day 1 of adulthood (see section 3.4.2). **(A)** Total load variation upon ageing; 3078 proteins were detected and quantified in the total fraction, for all the time points shown. **(B)** Aggregate load variation upon ageing; 965 proteins were detected and quantified in the insoluble fraction at all time points measured. **(C)** Correspondent total load variation upon ageing calculated only for the subset of proteins having a detected pellet fraction (965 proteins). Errors were calculated with a bootstrap method. Eppendorf pictures are shown for each bar charts for better visualisation of the quantity used for the analysis: Eppendorf in (A) and (C) imply the use of protein total abundance measurements, while the Eppendorf in (B) indicates the use of protein pellet fraction measurements. Bar charts are coloured according to the proteins involved in the calculation: black for the 3078 proteins, and red for the 965 proteins.

We estimated the total amount of intracellular protein material (see section 3.4.2) from the set of proteins detected in the total fraction at all time points (3078 proteins) (Figure 3.3A). We do not observe any significant change in the overall intracellular proteome amount upon ageing, even if about a third of proteins are found to change in abundance of at least 2-fold upon age [90]. Using the data of insoluble abundances, taking all proteins detected to form pellet in day 01 to day 17, we evaluated with the same procedure the whole intracellular aggregate amount with respect to day

01. Interestingly, we observe a sharp aggregate amount proliferation from day 12 of adulthood (Figure 3.3B), which is not reflected by an increase in the total amount of protein (Figure 3.3A). Also, this aggregate proliferation is not due to an increase in the total amount of proteins having a detected insoluble amount either, namely the supersaturated total amount. In fact, evaluating the total amount relative to day 01 for the 965 proteins forming aggregates throughout the days, we do not observe any change upon ageing (Figure 3.3C).

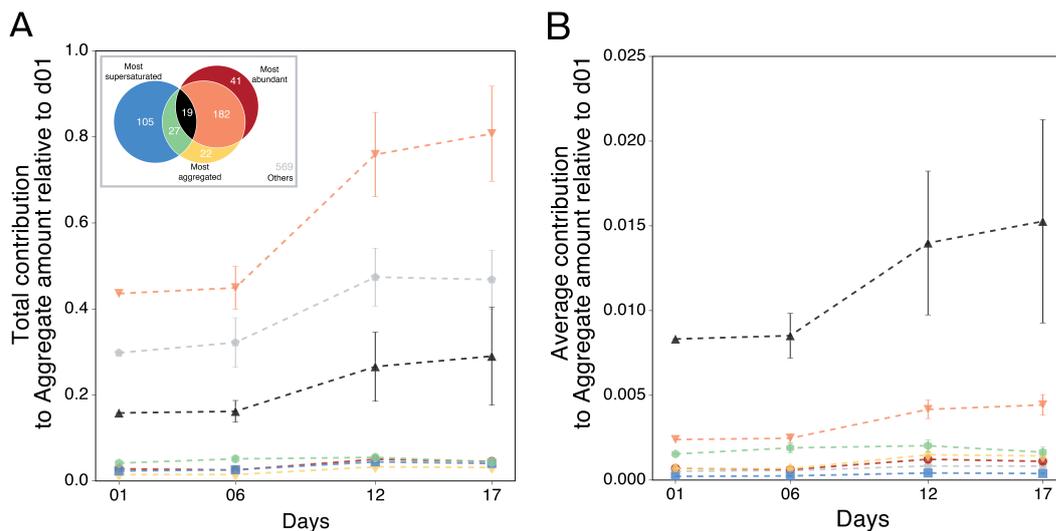
We have shown that the worm undergoes a reshaping of its proteome composition upon ageing without altering the total protein amount available in the cell. Even if the reshaping of the proteome does not involve a change in the total protein amount, it causes an overall change in the cellular environment that results into a boost of the total aggregate fraction. We estimated that between day 6 and day 12 the total aggregate load obtained by the contribution of each one of the 965 proteins detected with insoluble fraction is almost doubled (Figure 3.3B). Taken together, our results suggest that the mechanism by which proteins are further pushed far from the solubility edge with ageing is not due to an overall increase in their abundance content, but by some other mechanism, which reduces their solubility, which could be the disruption of a protein control machinery. Also, since the aggregates are increasing regardless of the total amount, this represents evidence that the critical concentration is not a general threshold that proteins might overcome, but it is a principle that is protein dependent, confirming the concept of the life on the edge.

3.2.4 Age-dependent aggregate proliferation in terms of supersaturation, aggregation and abundance

Having established that there is a proliferation of the insoluble protein amount starting from day 12 of adulthood, we now wish to probe which proteins and physical parameters are responsible for such behaviour. To this extent, we estimated the total (Figure 3.4A) and average (Figure 3.4B) contribution to the aggregate amount relative to day 01 considering all the six subsets of proteins we defined in Figure

3.2 from the overlaps of the 20% most supersaturated, abundant and aggregated proteins at day 12. A Venn diagram showing the screened mapping of the most supersaturated, abundant and aggregated proteins using the time-course insoluble fraction data of Figure 3.2, where intracellular proteins have been removed, is shown in Figure 3.4A. The total contribution to the aggregate amount relative to day 01 (Figure 3.4A, scatter plot) for each of the six resulting subsets was obtained by summing up the values calculated for the Pellet amount of Figure 3.3B only from the group of proteins belonging to the subset, whose number is indicated in the Venn diagram. The average contribution to the aggregate amount relative to day 01 (Figure 3.4B), was obtained from the total contribution of Figure 3.4A by normalising to the protein number in each subset.

Notably, proteins that are most supersaturated, abundant and aggregated on average contribute the most to the aggregate proliferation we detect in the old worm (black triangles, Figure 3.4B). Such proteins though are very few in numbers, hence they do not contribute much to the overall total aggregate amount (black triangles, Figure 3.4A). On the other hand, most abundant and most aggregated proteins (orange triangles in Figure 3.4), contribute the most the total aggregate amount and quite significantly to the average aggregate amount, despite their high number (182 proteins). Most interestingly, supersaturated proteins (blue squares, green circles, and black triangles in Figure 3.4) are not contributing to the aggregate proliferation unless they are abundant. Hence, taken together, our results indicate that upon ageing, supersaturation does not seem to play a role in the disruption of the solubility threshold. Since the present work is focused on the process of healthy ageing of the nematode, we do not exclude that supersaturation may play a key role in case or neurodegeneration [76, 131].

**Fig. 3.4**

Contribution of supersaturation, aggregation and abundance to aggregate amounts in ageing worms

Scatterplots of the contribution of the proteins belonging to the most supersaturated, aggregated and abundant proteins and corresponding overlaps to the variations in intracellular aggregate amounts upon ageing in wild-type worms of Figure 3.3B. Dashed lines are shown between points only as a guide to the eye. **(A)** Total pellet amount relative to day 01 of adulthood measured at different time points evaluated for the subset of intracellular proteins belonging to the classes of top 20% most supersaturated, most aggregated and most abundant and their intersections of Figure 3.2. Total contribution to the aggregate load for each subgroup is shown in a different colour and marker. Number of proteins in each subgroup and relative colour code are reported in the Venn diagram. **(B)** Average contribution, per protein, of to the aggregate load variation relative to day 01 for each class shown in (A). Errors were calculated with the bootstrap method used for Figure 3.3 (see section 3.4.2).

3.2.5 Functional protein classes responsible for the age-dependent aggregate increase

Finally, we identified the proteins most responsible for the increase in the aggregate levels upon ageing of Figure 3.3, and report the list in Table 3.1. These 32 proteins are the proteins that have the highest weight in the calculation of the change of aggregate amount relative to day 01, and which can account for the fold change in the aggregate levels. We found three major functional classes are associated with these proteins: molecular chaperones (small heat-shock proteins sip-1 and hsp-25 and heat shock proteins hsp-90 and hsp-70), RNA-binding and translation (ribosomal components and elongation factors); and structural activity (intermediate filaments, actin, tubulin).

Small heat shock proteins have been previously detected to co-aggregate and drive aggregation of proteins [241–245], in accordance with the hypothesis that *in vivo* aggregation may have a cytoprotective function by sequestering potentially toxic protein species [90, 245–247].

Ribosomal proteins and proteins related to translation functions have been previously observed to be significantly enriched in aggregate inclusions of older nematodes compared to younger nematodes [95], with implications on the lifespan of the organism. Proteins belonging to this functional class, also, have been predicted to be at the highest risk for oxidative destabilization, a suggested dominant source of protein stability and solubility loss upon ageing [248].

Table 3.1
List of proteins contributing the most to the aggregate proliferation

Protein ID, protein name, corresponding gene name and the protein family according to the Uniprot database [211] are shown for the 32 proteins found most responsible for the aggregate proliferation.

Protein ID	Protein names	Gene names	Protein families
Q20363	Stress-induced protein 1	sip-1 F43D9.4	Small heat shock protein (HSP20)
P53013	Elongation factor 1-alpha (EF-1-alpha)	eft-3 F31E3.5; eft-4 R03G5.1	TRAFAC class translation factor GTPase superfamily, Classic translation factor GTPase, EF-Tu/EF-1A subfamily
Q19286	Intermediate filament protein ifb-2 (Cel IF B2) (Intermediate filament protein B2) (IF-B2)	ifb-2 F10C1.7	Intermediate filament
Q18688	Heat shock protein 90 (Abnormal dauer formation protein 21)	daf-21 C47E8.5	Heat shock protein 90
Q19289	Intermediate filament protein ifb-1 (Cel IF B1) (Intermediate filament protein B1) (IF-B1)	ifb-1 F10C1.2	Intermediate filament
Q45EJ8	Lin-5 (Five) Interacting protein	lfi-1 CELE_ZC8.4 ZC8.4	
Q21067	Intermediate filament protein ifc-2 (Cel IF C2) (Intermediate filament protein C2) (IF-C2)	ifc-2 M6.1	Intermediate filament
P29691	Elongation factor 2 (EF-2)	eef-2 F25H5.4	TRAFAC class translation factor GTPase superfamily, Classic translation factor GTPase, EF-G/EF-2 subfamily
Q9TY23	Prion-like-(Q/N-rich)-domain-bearing protein	pqn-22 C46G7.4 CELE_C46G7.4	
O02056	60S ribosomal protein L4	rpl-4 B0041.4	Ribosomal protein L4P
Q5H9M9	Heat Shock Protein	hsp-25 C09B8.6 CELE_C09B8.6	Small heat shock protein (HSP20)
P27604	Adenosylhomocysteinase (AdoHcyase) (EC 3.3.1.1) (Protein dumpy-14) (S-adenosyl-L-homocysteine hydrolase)	ahcy-1 ahh dpy-14 K02F2.2	Adenosylhomocysteinase

Continued on Next Page...

Protein ID	Protein names	Gene names	Protein families
P09446	Heat shock 70 kDa protein A	hsp-1 hsp70a F26D10.3	Heat shock protein 70
O45815	ACTin	act-5 CELE_T25C8.2 T25C8.2	Actin
Q9XW17	Cytokinesis, Apoptosis, RNA-associated	car-1 CELE_Y18D10A.17 Y18D10A.17	
Q93572	60S acidic ribosomal protein P0	rpa-0 F25H2.10	Ribosomal protein L10P
Q20206	Ribosomal Protein, Small subunit	rps-11 CELE_F40F11.1 F40F11.1	Ribosomal protein S17P
Q27389	60S ribosomal protein L13a	rpl-16 M01F1.2	Ribosomal protein L13P
P04255	Histone H2B 1	his-11 ZK131.5; his-15 ZK131.9; his-29 F35H10.11; his-34 F17E9.9; his-44 F08G2.1	Histone H2B
G5EFP2	Uncharacterized protein	CELE_ZK1321.4 ZK1321.4	
P90900	Intermediate filament protein ifa-4 (Cel IF A4) (Intermediate filament protein A4) (IF-A4)	ifa-4 K05B2.3	Intermediate filament
O17921	TuBulin, Beta	tbb-1 CELE_K01G5.7 K01G5.7	Tubulin
Q9BKU5	Uncharacterized protein	CELE_Y37E3.8 Y37E3.8	Ribosomal protein L15P
P10984	Actin-2	act-2 T04C12.5	Actin
Q93573	Translationally-controlled tumor protein homolog	tct-1 F25H2.11	TCTP
Q27535	Probable arginine kinase ZC434.8 (AK) (EC 2.7.3.3)	ZC434.8	ATP:guanido phosphotransferase
B3WFT8	Uncharacterized protein	C14F11.4 CELE_C14F11.4	
P49405	60S ribosomal protein L5	rpl-5 F54C9.5	Ribosomal protein L18P
O17687	NASP (Human Nuclear Autoantigenic Sperm Protein) homolog	nasp-2 C50B6.2 CELE_C50B6.2	

Continued on Next Page...

Protein ID	Protein names	Gene names	Protein families
P91917	Obg-like ATPase 1	ola-1 tag-210 W08E3.3	TRAFAC class OBG-HflX-like GTPase superfamily, OBG GTPase, YchF/OLA1 subfamily
P54412	Probable elongation factor 1-gamma (EF-1-gamma) (eEF-1B gamma)	eef-1G F17C11.9	
Q20751	Eukaryotic translation initiation factor 5A-2 (eIF-5A-2) (Initiation factor five protein 2)	iff-2 F54C9.1	EIF-5A

Notably, the protein contributing the most to the aggregate proliferation is sip-1, a small heat shock protein that becomes active under acidic conditions and is essential for nematode development and reproduction [249]. Interestingly, RNA binding proteins and cytoskeletal proteins have been identified as more specific and prominent substrates of sip-1 with respect to other molecular chaperones [249]. Cytoskeletal proteins, indeed, have also been suggested to be specific substrates of the Hsp20 molecular chaperone family [250].

3.3 Conclusions

We find first evidence of the general validity of the life on the edge hypothesis in the adult wild-type nematode *C. elegans*, and show that supersaturated abundant and aggregated proteins escape the evolutionary pressure. Also, we found that the way proteins exceed the solubility edge reflects a sudden proliferation in the aggregate levels which is independent of the intracellular total content, and which is not due to supersaturation. Proteins most responsible for such an increase can be mapped into three main functional classes, which have been previously associated with the protein homeostasis machinery: structural proteins, proteins involved in translation, and heat shock and small heat shock proteins. The findings that we have reported here, therefore, suggest that critical concentration is a principle that is protein-specific and not a general threshold which proteins might overcome to aggregate, and that

the way proteins cross the edge is not by an increase in their overall expression levels, but by decreasing their solubility due to some other cellular factors, such as disruptions in the protein control machinery.

3.4 Materials and Methods

3.4.1 Calculations of total, soluble and pellet abundances in the adult nematode

Data of protein abundances (absolute LFQ values) from proteomic-wide mass spectrometry measurements of WT strain nematode *C. elegans* were obtained from Table S1F of Ref. [90]. The data contained up to three replicas of supernatant, pellet and total fraction measurements per protein taken at day 12 of the nematode life with respect to the exit of the L4 larval stage. For each type of measurement (either total, supernatant or pellet) only proteins found in at least two replicas were considered in the calculations, and average and standard error were evaluated.

To be consistent with the definition of protein solubility considered as the maximum soluble fraction in present of a non-negligible amount of pellet, we defined the total (T), soluble (S) and pellet (P) abundance as follows

$$T = \frac{2t_{av} + s_{av} + p_{av}}{3} \quad (3.3)$$

$$S = \frac{2s_{av} + t_{av} - p_{av}}{3} \quad (3.4)$$

$$P = \frac{2p_{av} + t_{av} - s_{av}}{3} \quad (3.5)$$

where t_{av} , s_{av} , p_{av} are the total average, supernatant average and pellet average of the mass spectrometry LFQ values respectively. Proteins resulting with negative S

and P values were excluded in the calculation. With this normalisation, no preference is given to any specific quantity detected by mass spectrometry, and the Eq. 3.6

$$T = S + P \quad (3.6)$$

holds true. Errors on T , S and P were obtained using the law of error propagation of the independent variables t_{av} , s_{av} and p_{av} . Only proteins satisfying the condition of having a resulting percentage error not over 20% in all quantities were retained in the analysis. This resulted in 1828 proteins, whose total abundance (T) and soluble abundance (S) are reported in \log_{10} scale in Figure 3.1B with red filling markers. Pellet abundance (P) and soluble abundance (S) of these 1828 proteins are plotted in \log_{10} scale in Figure 3.2. Also, we evaluated T and S for protein not having a detectable pellet fraction, i.e. having $p_{av} = 0$. Using the same percent error threshold ($\leq 20\%$) for the sake of signal reproducibility, we show the 674 proteins not having aggregates in Figure 3.1B (grey points).

3.4.2 Calculation of total and aggregate load variation upon ageing

Data of protein absolute and relative abundances (absolute LFQ values and SILAC values) for total and aggregate protein measurements in wild-type strain worms at day 1, 6, 12 and 17 of adulthood were retrieved from Tables S1B,E of Ref. [90]. Amounts relative to day 1 of adulthood reported in Figure 3.3 for total (panel A), total having pellet (panel B) and pellet (panel C) were calculated with the same procedure, described as follows.

Only proteins detected and quantified in all the days and for both SILAC and LFQ datasets were considered in the calculations. This resulted into 3694 proteins with total measurements, of which 1083 also reporting pellet measurements. This common ensemble was then filtered to remove proteins known, or predicted, to be

extracellular. We relied on both the UniProt classification of extracellular proteins (Subcellular location entry) [251], the outcome of the Signal P predictor of signal peptides [252] and the WoLF PSORT [253] predictor of subcellular locations for the removal of all possible extracellular proteins.

The remaining intracellular ensemble consisted in 3078 proteins with detected total abundance at all time points (used for Figure 3.3A) and a subset of 965 having also a detected pellet fraction (used for Figure 3.3B,C). For each protein, two quantities were estimated: the abundance weight W of the protein p_n to the whole ensemble at a given time point d_i , defined as the ratio of the absolute LFQ value of the protein and the sum of all the proteins absolute LFQs at the given d_i day

$$W(p_n)|_{d_i} = \frac{LFQ(p_n)|_{d_i}}{\sum_{n=1}^N LFQ(p_n)|_{d_i}} \quad (3.7)$$

where N is the number of proteins in the ensemble; and the abundance change of the protein $\Delta A(p_n)|_{d_i}^{d_j}$ between this given time point and any other time point measured (d_j). For calculation of $\Delta A(p_n)|_{d_i}^{d_j}$, we relied on the SILAC labelled data

$$\Delta A(p_n)|_{d_i}^{d_j} = \frac{SILAC(p_n)|_{d_j}}{SILAC(p_n)|_{d_i}} \quad (3.8)$$

Protein loads change ΔL between time point d_j , and time point d_i , are obtained by summing the contribution from each protein of the weighted abundance increase obtained by combining $W(p_n)|_{d_i}$ and $\Delta A(p_n)|_{d_i}^{d_j}$

$$\Delta L|_{d_i}^{d_j} = \sum_{n=1}^N W(p_n)|_{d_i} \cdot \Delta A(p_n)|_{d_i}^{d_j} \quad (3.9)$$

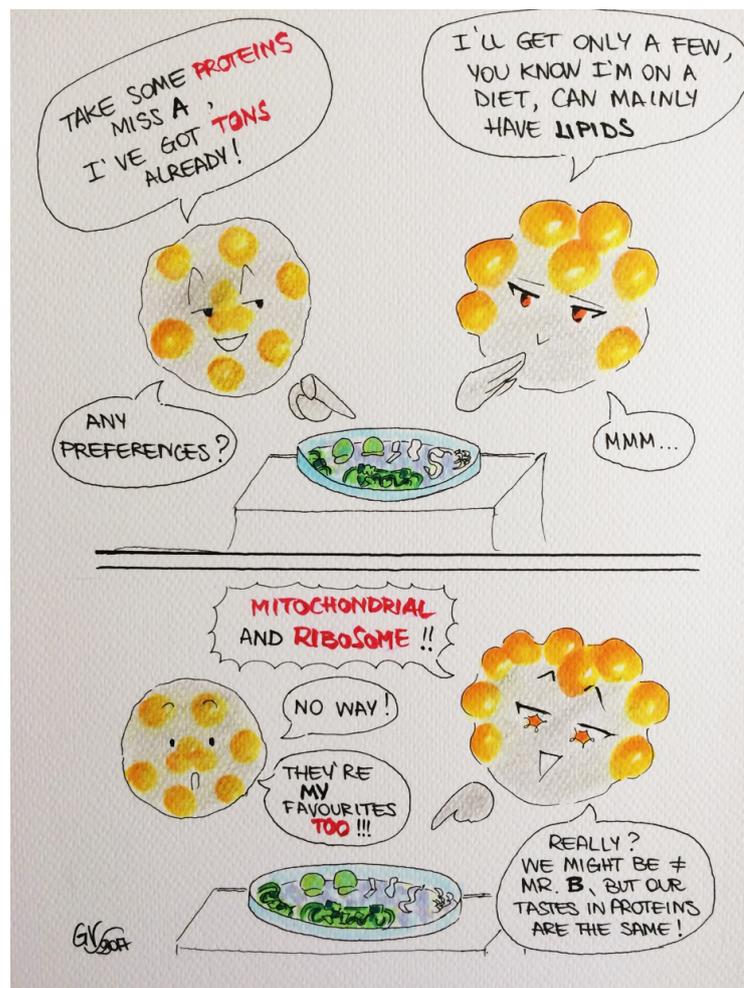
To obtain loads variation relative to day 1 of adulthood without imposing a preferential bias on the absolute LFQ data measured at day 1, load variations relative to any day detected were normalised to the correspondent value for day 1 and subsequently averaged

$$\text{Load change}_{d_{01}}^{d_j} = \frac{\sum_{n=1}^4 \frac{\Delta L_{d_i}^{d_j}}{\Delta L_{d_i}^{d_{01}}}}{4} \quad (3.10)$$

Using Eq. 3.10 total abundance load change upon age with respect to day 1 was calculated for the detected intracellular proteome (3078 proteins) and is reported in Figure 3.3A. Total amounts in presence of pellet and pellet amounts changes were also calculated from the subset of 965 proteins, and are reported in Figure 3.3B,C. Errors on each time point reported in Figure 3.3 were obtained using a bootstrap method on the weighted abundance change ensemble.

Chapter 4

The interactome of misfolded protein oligomers



4.1 Introduction

Misfolded oligomers formed at the early stage of the aggregation process are thought to play a central role in the onset and progression of neurodegenerative diseases, as they have been found to be the most toxic aggregated species to the cell [62, 98, 115, 146–151] (see Chapter 1, section 1.2.5). It is still unclear how these oligomers cause the cell damage and their contribution to neurodegeneration. Experiments, aimed at revealing the molecular mechanism of toxicity of the oligomers, have shown that:

- these aggregates can interact with and disrupt membranes [149, 150, 254–257];
- they are able to recruit and affect the functionality of proteins [82, 258];
- they can form pores on the cell membrane causing an uncontrolled exchange of compounds between the intracellular and extracellular space [257, 259–262].

The mechanism of toxicity that involves the interaction with the lipid membrane has been observed for oligomers formed with the HypF-N protein [150]. Even if this protein is not associated with any disease, it can form amyloid-like aggregates which have been shown to mimic the synaptotoxicity of $A\beta$ aggregates in Alzheimer's disease [152]. HypF-N oligomers, when placed outside cells, are found to interact with the lipid membrane [149, 150]. The cell viability is measured by quantifying the damage upon disruption of the membrane, that leads to a calcium intake and ultimately to apoptosis [149, 150]. Upon given conditions two types of HypF-N oligomers can be formed, which are stable and have been well characterised in structure. These two types of oligomers (type A and type B, see Figure 4.1) are structurally different and result into a very different biological activity in terms of interaction with the lipid membrane. Oligomers that expose to the surface hydrophobic and structurally disorganised patches are able to penetrate the membrane, and cause an influx of calcium ions that can lead to apoptosis, while oligomers characterised by buried and structured hydrophobic regions cannot readily permeate the membrane and remain attached to the outer surface [150].

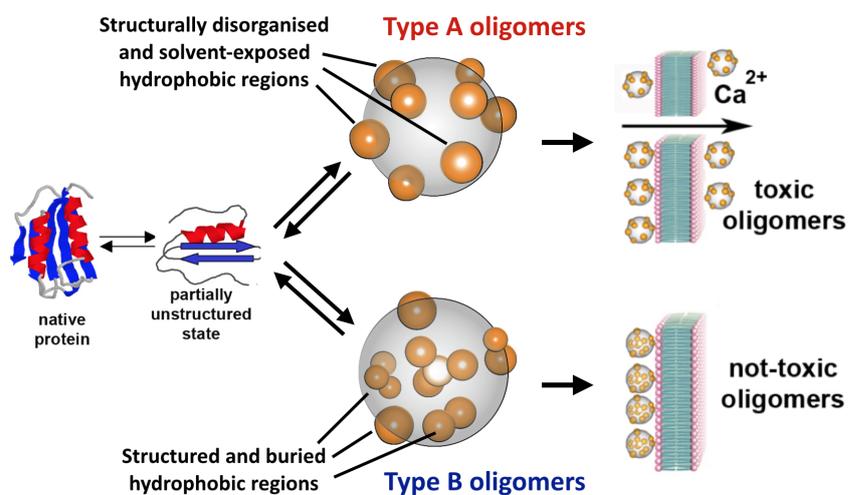


Fig. 4.1

Schematic representation of the two structurally different HypF-N oligomer types

The known mechanism of toxicity is mediated by the interaction with the lipid membrane cellular component. Type A oligomers, which show structurally disorganised and solvent-exposed hydrophobic regions are able to penetrate the membrane, while type B oligomers, which are more structured and have hydrophobic patches buried in the core, cannot.

Adapted from [150].

While the interaction of extracellular oligomers with the membrane of cells has been studied quite extensively, little is known about what happens when oligomers are formed intracellularly, like the case of α -synuclein and tau, or when they penetrate into the cells. Also, a recent study reported the existence of oligomeric species that induce membrane leakage, but are not cytotoxic, showing hence the absence of one-to-one relationship between the induction of cellular toxicity and the disruption of membranes [263]. Given more than half of the cell dry mass is constituted by proteins [167], characterising the interaction of the oligomers with the proteome cellular component could provide a framework for discovering an alternative mechanism of toxicity which could be relevant in the neurodegeneration process. The aim of this work was therefore to identify protein binders of oligomers, by taking advantage of the two structurally different type A and type B HypF-N oligomers. To address this study, we used MS-based proteomics and bioinformatic analyses to characterise the

interactome of misfolded oligomers (Figure 4.2), i.e. the pool of proteins that can bind the oligomers.

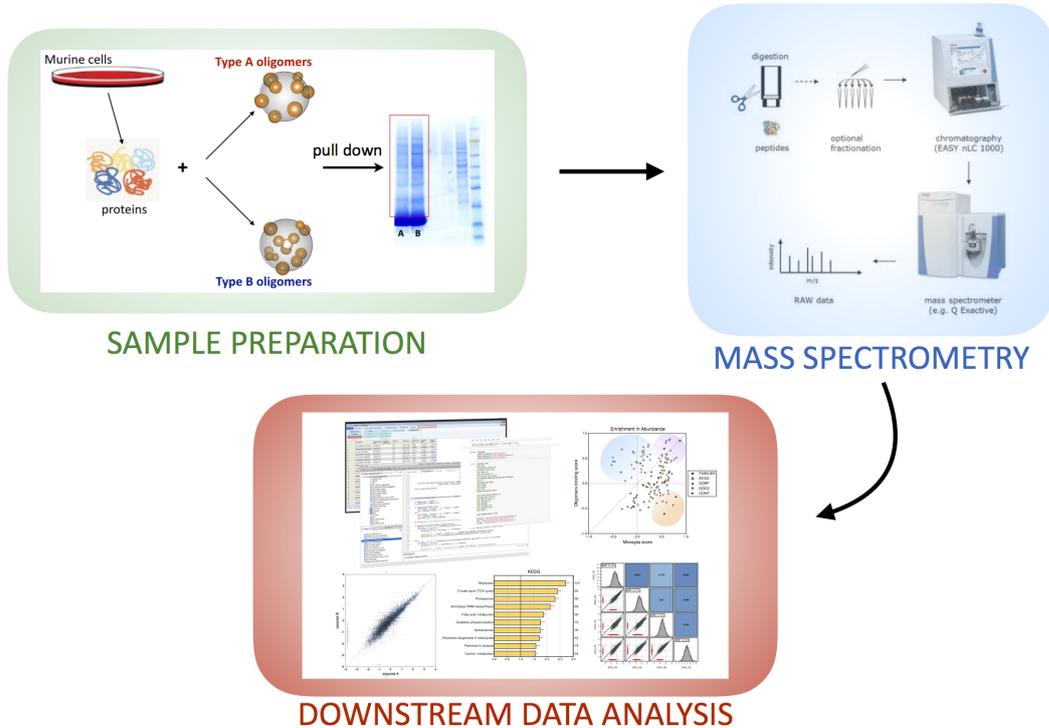


Fig. 4.2

MS-based proteomic schematic workflow

Cultured murine microglia cells were lysed to extract the proteome which was then subject to the presence of either type of oligomers. Proteins that bound to the oligomers in the pull-down assay were then analysed by Liquid Chromatography Tandem-mass Spectrometry (LC-MS/MS). Four biological replicas were performed for each type of oligomer interactome. Resulting MS raw data were processed with the MaxQuant software for quantitative proteomics, for identification and quantification of the proteins in the samples. Processed data were subject to data analysis and complemented with bioinformatics functional annotation analysis.

The two interactomes of type A and type B oligomers were obtained by using quantitative proteomics. We find that the structural difference of the two oligomers leads to a great difference in the degree of binding to proteins. The oligomers with a small number of hydrophobic-exposed patches, type B, bind proteins much more strongly than type A. The number of proteins that are pulled down with the oligomers accounts for more than 2000 entries. Interestingly, the two pools of proteins are basically the same, meaning that the structural difference of the oligomers

is not responsible for binding specificity. Among the proteins found to interact with the oligomers, we detect a significantly high enrichment in mitochondrial proteins, molecular chaperones, ribosomal components and RNA binding proteins. Interestingly, some of the proteins falling in these categories have previously been reported to co-aggregate and be enriched in aggregate inclusions during ageing [95, 241–245].

4.2 Results

4.2.1 Reproducibility of MS data

To study the ability of type A and type B oligomers to interact with the protein component of the cell, we extracted proteins from N13 murine culture cells and we incubated them *in vitro* in the absence or in the presence of type A and type B oligomers. After the incubation each sample was centrifuged to pull down the oligomers and the resulting pellet fraction were subjected to MS-based proteomics (see Materials and Methods, section 4.4). In order to ensure statistical significance of the results, four biological replicas were performed. The resulting MS raw data from all the replicas in the two conditions were processed with the MaxQuant software (see Materials and Methods, section 4.4.2) and subsequent protein identities and abundance data (iBAQ values in log₁₀ scale, see Material and Methods section 4.4.2) were retrieved.

To verify the consistency of the interactomes and the biological variability, we report in Figure 4.3 the matrix of 6 correlation plots obtained comparing the iBAQ intensities (in log₁₀) of the 4 biological replicas per oligomer type (Figure 4.3A for type A oligomers and Figure 4.3B for type B oligomers respectively). In each scatterplot of two compared replicas, green points represent abundance data of proteins detected in both replicas, while red points at the corner of the plot indicate abundances of proteins that have been detected only in one of the two replicas considered. We observe very high correlations between the replicas for both conditions, with Pearson's coefficient of correlation ranging between 0.73 and

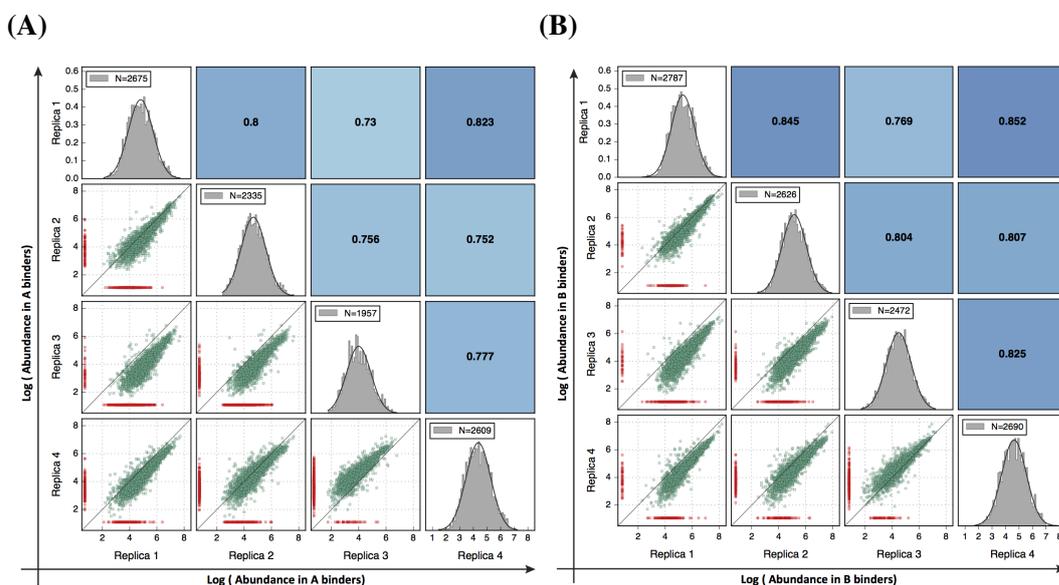


Fig. 4.3

Reproducibility of the MS data of the biological replicas

The matrix of 6 correlation plots between the iBAQ intensities of the 4 biological replicas, for both oligomers types is shown. For each scatterplot (bottom left triangle of the matrix) green points represent proteins that have been detected in both replicas, while red points at the corners of the plot represent proteins that have been detected only in one of the two considered replicas, and hence possessing only one abundance value. Abundances values are iBAQ intensities in log₁₀ units.

In the upper right triangle of the matrix, Pearson correlation coefficients of the corresponding transposed scatterplots are shown, with colour code (from red to blue) following the indicated values of correlation coefficient.

Along the diagonal, histogram of the log₁₀ iBAQ abundances distribution of each replicas are shown in grey with a gaussian fit line in black. Number of proteins detected in the replica (N) are indicated in the top left corner of each distribution plot. **(A)** Correlation matrix for the 4 biological replicas of the proteins interacting with type A oligomers. **(B)** Correlation matrix for the 4 biological replicas of the proteins interacting with type B oligomers.

0.82 for the proteins found interacting in the sample with type A oligomers (Figure 4.3A), and ranging between 0.76 and 0.85 for the proteins pulled down with type B oligomers (Figure 4.3B).

The results show that measurements of the abundances of the protein pulled down with the oligomers are very reproducible and slightly more consistent against B oligomers, mainly due to the presence of a stronger MS signal given by the higher abundances of proteins detected after the incubation with with type B oligomers (discussed in section 4.2.2). Also, more proteins have been detected in each replica to bind type B oligomers than type A, as we can see from the number of proteins shown in the distribution histogram of abundances of the correlation plots: 1957 to 2675 proteins detected with type A, (Figure 4.3A) and 2472 to 2787 proteins detected with type B (Figure 4.3B).

4.2.2 The structural difference of the oligomers modulates the binding strength to proteins

To get insights into the ability of the two different oligomers to bind proteins, intensity-based absolute quantification (iBAQ) values of protein abundances (see Materials and Methods, section 4.4.2) of type A and type B binders were compared using MaxQuant. The iBAQ value is proportional to the amount of the protein in the sample. This amount is the result of the contribution of two independent quantities: the physiological abundance of the protein in the lysate, and the binding affinity of the protein to the oligomer. Therefore, the iBAQ value can be considered a function of the physiological abundance and the binding affinity to the oligomer.

In each biological replica, the physiological abundance of the protein in the lysate is approximately conserved between the two different oligomeric samples, since the oligomers were incubated with the same batch of extracted proteins. Hence, in a given biological replica the comparison, for any protein binder, of the iBAQ values of type B oligomers versus type A oligomers represent the difference in binding that the oligomers exert on that protein.

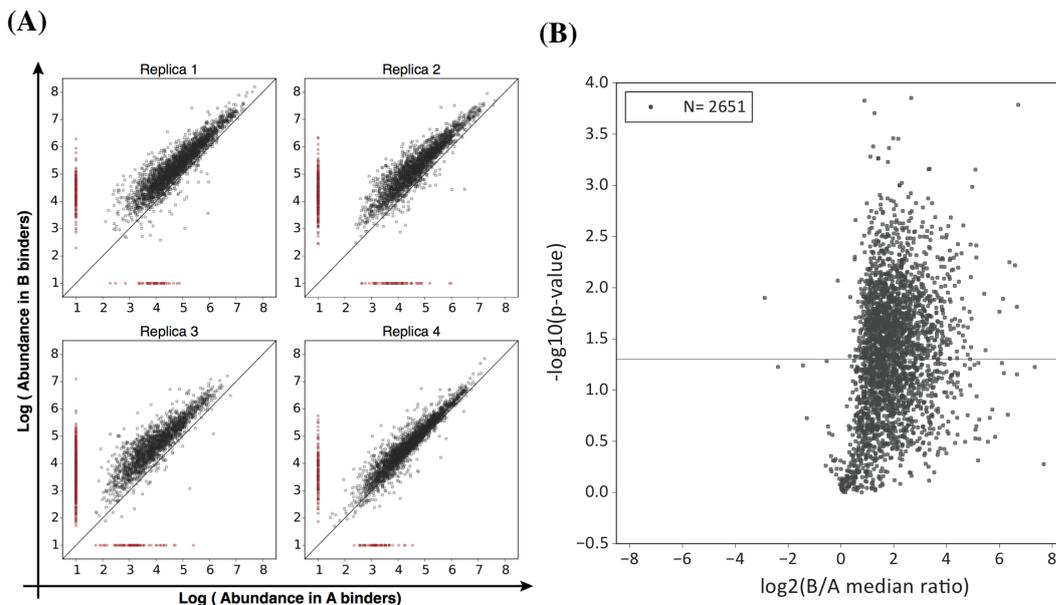


Fig. 4.4
Difference in binding power of type A and type B oligomers

(A) Scatterplot matrix of the replica-specific pairwise comparison of \log_{10} iBAQ values of proteins bound to B oligomers versus those bound to A oligomers. Black points in each scatterplot represents proteins quantified in the biological replica both in the sample of A interactors and in that of B interactors. Red points represent proteins that have been detected only in one condition (either B bound or A bound) in that specific replica. The presence of the greatest fractions of proteins above the bisector line in all 4 replicas shows that B oligomers have much higher capacity in binding proteins than do have A oligomers. (B) Volcano plot showing the average of the ratios of B binders versus A binders taken per biological replica, in \log_2 , and their significance (see Material and Methods, section 4.4.2 for evaluation). The 2651 points are the number of proteins detected at least twice in the corresponding sample of A binders and B binders.

Figure 4.4A show the scatterplot matrix of the iBAQ values (in log10) of B binders versus those of A binders per biological replica. Each black point represent a protein quantified by MaxQuant to be in the sample of both oligomers (after the incubation), in that biological replica. Red points are abundances detected only for one condition, either binding to A oligomers (horizontal points) or binding B oligomers (vertical points) again in that replica. Since nearly all points lie over the bisector line (black line in each plot) in all replicas, type B oligomers bind all proteins much more strongly than do type A oligomers. Therefore, the structural difference of the two oligomers induces a different degree of binding to the proteome cellular component.

This difference in binding affinity becomes even more clear in the volcano plot shown in Figure 4.4B. Each black point in the plot is a protein which has been detected to bind both B oligomers and A oligomers in at least two corresponding biological replicas. On the x axis, the averaged ratio in log2 of the intensities from iBAQ values of B binders over A binders is shown, and its corresponding significance is plotted on the y axis (see section 4.4.2 for evaluation). Positive values on the x axis indicate proteins able to bind more to B oligomers than A oligomers, while negative values implies that these proteins bind stronger A oligomers. The clear cut asymmetry of the plot towards positive values means that the cellular component represented by proteins have much stronger affinity to bind B oligomers, which have more structured and buried hydrophobic patches. Hence, structurally different oligomers can lead to a different biological activity in terms of the interaction with the proteome cellular component and the structural differences in the oligomers can be responsible of the binding affinity to proteins.

4.2.3 The two types of oligomers interact with similar pools of proteins

To shed light on the proteins contained in the interactomes of each oligomer, we combined the results of all biological replicas per condition by estimating the median

abundance fraction that each binder protein contributes to with respect to all the binders found in the sample (see Materials and Methods, section 4.4.2). The median abundance fraction represents a measure of the abundance weight that each protein has with respect to the total mass of proteins that have been found to bind the concentration of oligomers.

We define as shared binders between the two types of oligomers proteins detected in at least three biological replicas in both conditions. Complementarily, we define as specific binders proteins that have been detected at least thrice in one condition, but have never been detected in the other.

We found that 2280 proteins are common binders of the two types of oligomers (blue points, Figure 4.5), while only 18 proteins are found only in one condition (17 bound to type B oligomers and 1 bound to type A oligomers, red points in Figure 4.5). 17 of these 18 proteins, however, have been detected in the lowest range of abundance values (\log_{10} percentages in abundances $\ll -2$), hence we cannot exclude the possibility that the signal in the other condition was present but:

1. resulted lower than the detection threshold of the mass spectrometer
2. was discarded due to the partial stochasticity of the peptides quantification in the MS process, in which only the top 20 peptides that are eluting in a given time frame are kept for MS/MS analysis (see Materials and Methods, section 4.4.1)

Therefore, the abundance percentage of these 17 specific binders is statistically not high enough for them to be considered exclusive binders of the given oligomer type.

Only one protein that binds only to type B oligomers is found in the upper range of the abundance fraction, more than one standard deviation above the average logged-abundance of the ensemble of B binders, and can be considered a specific binder. This protein, SET, which accounts for 0.066% of the total abundance of proteins in the sample (see Figure 4.5, highest red point in the vertical stack), is a

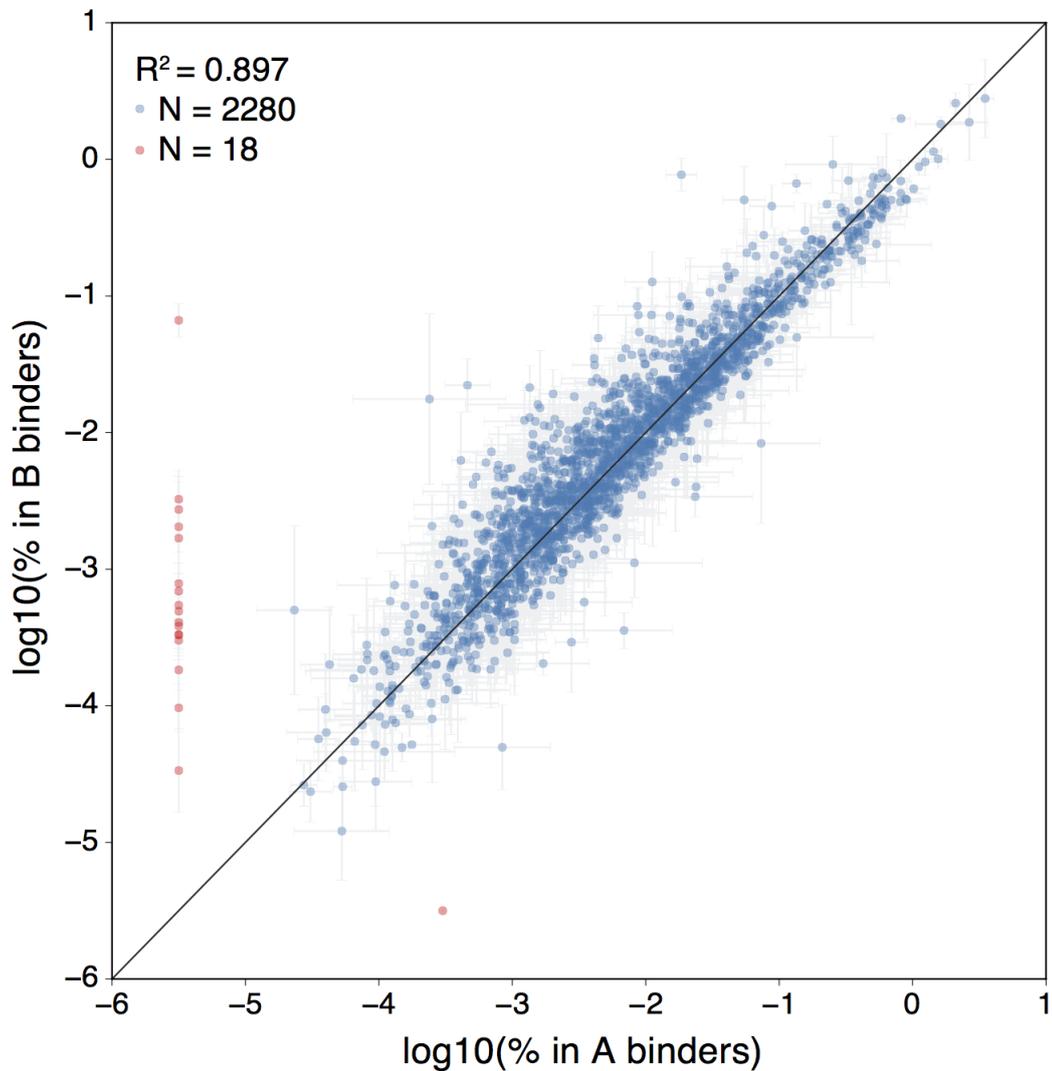


Fig. 4.5
Oligomers bind the same pool of microglia proteins

Scatterplot of the median percentage in abundance of proteins binding type B oligomers versus protein binding type A oligomers (in \log_{10}). Blue points in the plot represent proteins detected binding both oligomers in at least three replicas out of four. Red points at the corners are proteins that have been detected at least three times bound to one oligomer type and never detected to bind to the other oligomer type. Errors are median absolute deviations (see Materials and Methods, section 4.4.2 for calculations). Abundance percentage data span about 6 orders of magnitude and are highly correlated in the two conditions (binding to A oligomers and to B oligomers). The Pearson's correlation coefficient for the shared binders of the two types of oligomers is indicated on the upper left of the scatterplot.

multitasking protein which has been reported to be involved in transcription, histone chaperoning, nucleosome assembly and apoptosis [211, 264].

Given that more than 99% of the proteins detected are binding both the two types of oligomers types and that the signal of specific interactors is not statistically significant, we can conclude that despite their different binding strength, the two structurally different oligomers bind the same pool of proteins, therefore defining a single interactome. The number of interactors from this *in vitro* experiment is more than 2200 and spans about six orders of magnitude in the ensemble of abundance fractions, ranging from $\sim 10^{-5}\%$ up to $\sim 2.7\%$ abundance weight in the oligomer samples (0.44 in log10 scale, Figure 4.5).

Also, Figure 4.5 shows that the relative abundances of binders are conserved in the two conditions, since the correlation between the abundance percentages in the two conditions is about 0.897. Such strong correlation implies that, regardless of the overall intensity of binding, proteins bind the two structurally different oligomers in the same proportions. The different relative abundance among the proteins in the sample is therefore independent of the structure of the oligomer. The type of oligomer influences the total amount of protein that bind the oligomer, but not the relative abundances within the oligomers interactome neither the specificity of the proteins. In terms of relative abundances, we can conclude that the two oligomers bind similar proteins in similar proportions and hence result into a similar interactome.

In the next sections, we characterise the biological features of the constituents of the interactome. We used relative abundance data from the sample of type B oligomers as reference for the interactome as they are more accurate (see correlations in Figure 4.3 and section 4.2.1), but the results of the biological features (e.g. biological processes, pathways and functions enriched) are to be considered general for the two types of oligomers (see Material and Methods, section 4.4.3).

4.2.4 The biological constituents of the misfolded oligomers interactome

Next we characterised the interactome of the misfolded oligomers, by calculating the enrichment of protein groups belonging to specific biological categories with respect to the reference mouse proteome (23619 proteins) used in the processing of MS data (see Materials and Methods, section 4.4.3). Annotation terms in protein families, KEGG pathways and GO categories that resulted significantly enriched are plotted in Figure 4.6. We find that RNA binding proteins (pink coloured bars in Figure 4.6) appear in many annotation terms and are the most overrepresented in the oligomers interactome. In this class, the strongest signal comes from ribosomal components or translation-related proteins, which have in most annotations a relative maximum enrichment (RME, see Materials and Methods, section 4.4.3) higher than 0.6, an enrichment ratio (ER) bigger than 4 and corrected p-values smaller than 10^{-4} . Class I and II aminoacyl-tRNA synthetase, splicing factor SR and the spliceosome terms are also greatly enriched in this class.

Proteins belonging to the protein homeostasis network (orange coloured bars in Figure 4.6) are also at the top of the significantly highly enriched terms. In terms of protein families, we find all components of the peptidase T1A family and nearly all of the peptidase T1B (RME=0.82, $p < 10^{-4}$), together with the 14-3-3, the TCP-1 chaperonin and the Hsp 70 family (RME of 0.86, 0.73 and 0.44 respectively, with p-values $p < 10^{-3}$, $p < 10^{-4}$ and $p < 0.01$). In terms of KEGG pathways, the strongest signal comes from the proteasome and protein export pathways, both highly significant.

Another group of enriched annotation categories that are among the most overrepresented in the interactome is constituted by mitochondrial proteins and protein involved with energy metabolism (green bars in Figure 4.6). Indeed, the mitochondrion is the cellular component that is mostly enriched after the ribosome, with an enrichment ratio of 3 and a pvalue $p < 10^{-4}$. The citrate cycle, fatty acid metabolism and oxidative phosphorylation are the most relevant pathways overrepresented for

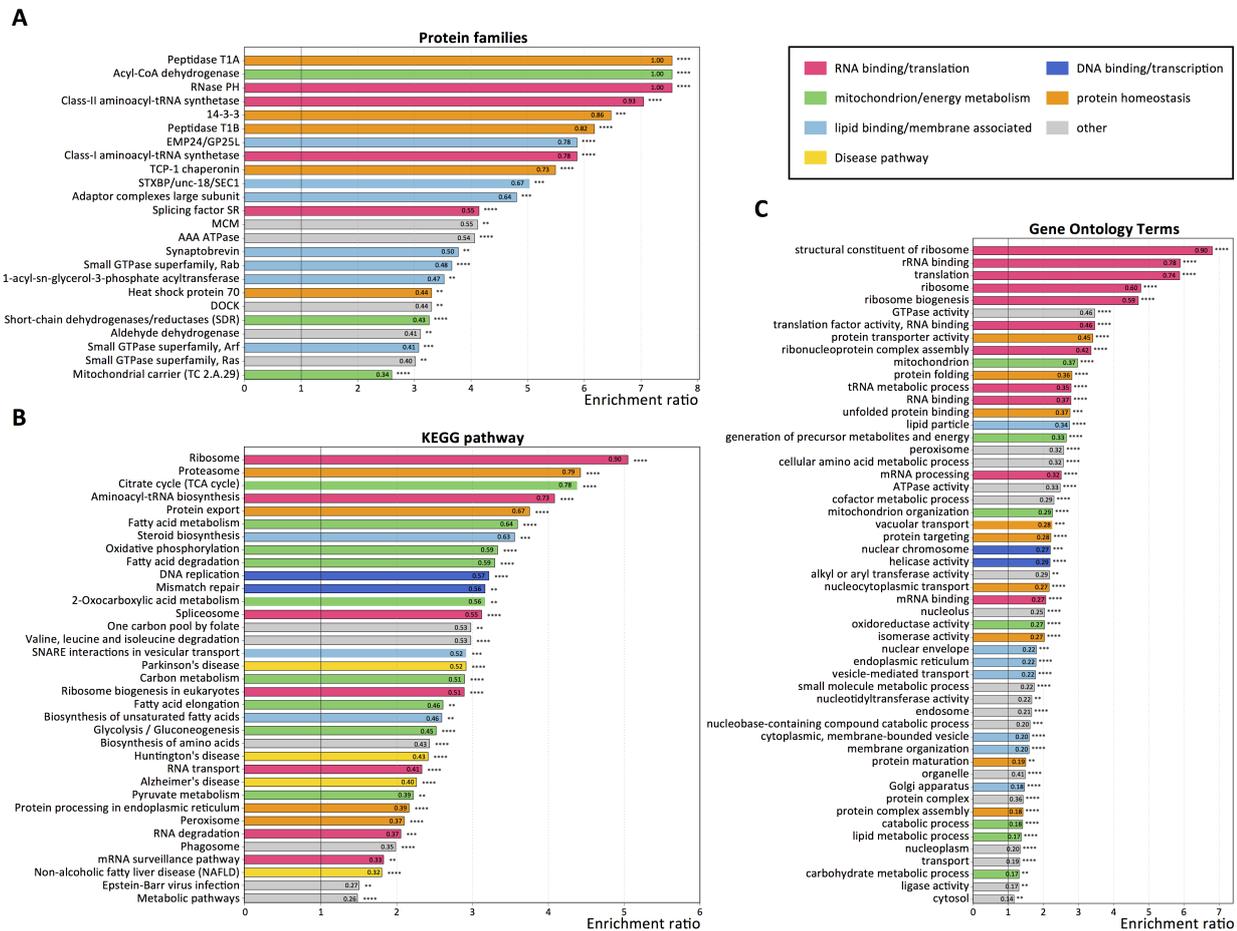


Fig. 4.6

Functional enrichment of the proteins in the oligomers interactome

Enrichment of protein families, KEGG pathways and GO categories of proteins in the oligomers interactome. The interacting proteins were grouped and categorised according to their functional annotations, see section 4.4.3. Highly significant ($p < 0.01$) overrepresented annotation terms are bar-plotted according to their enrichment ratio (ER). Values of relative maximum enrichment (RME) (see Materials and Methods, section 4.4.3) are shown inside the bars. Stars next to the bars indicate the significance of the enrichment, calculated with the Fisher exact test and corrected with the Benjamin- with the following scheme: **= $p < 0.01$,

= $p < 10^{-3}$, *= $p < 10^{-4}$.

this class in the interactome. Also, all the members of the Acyl-CoA dehydrogenase family have been detected to bind the oligomers, hence having an ER>0.75 and a RME of 1 (with $p < 10^{-4}$).

Notably, among the KEGG pathways enriched in the interactome, we find the occurrence with high significance of Parkinson's, Alzheimer's and Huntington diseases, the most important neurodegenerative disorders (yellow bars in Figure 4.6).

4.2.5 Mitochondrial and ribosomal proteins are the preferential binders of misfolded oligomers

Figure 4.6 and Figure Figure 4.5 show which are the main biological components of the oligomers interactome and what are the different abundance fractions within the single binders respectively.

In order to understand if the difference in abundance between two given binders is due to a different binding affinity or just to a difference in their natural abundance in the cell, we compared the abundances of the binders in the interactome with their natural abundances in microglia obtained using proteome-wide MS abundance data of mouse microglia cells recently published by Sharma and co-workers [207]. Proteins IDs of the interactome sets were mapped and matched to the available set of data of the proteome of microglia cells. iBAQ values in log10 scale (to ensure a normal distribution) from both sets were zscored, resulting into abundances distributions shifted to have mean value centered in 0 and standard deviation equal to 1. With this transformation, the distributions of protein abundances in the two sets were directly comparable. Figure 4.7 shows the result of this comparison. We do not observe a correlation between the two abundances, which suggests that misfolded oligomers have some preferential binders, to which they interact with more affinity. Also, protein cellular levels are not sufficient to reproduce the different abundance of the protein binders in the interactome sample.

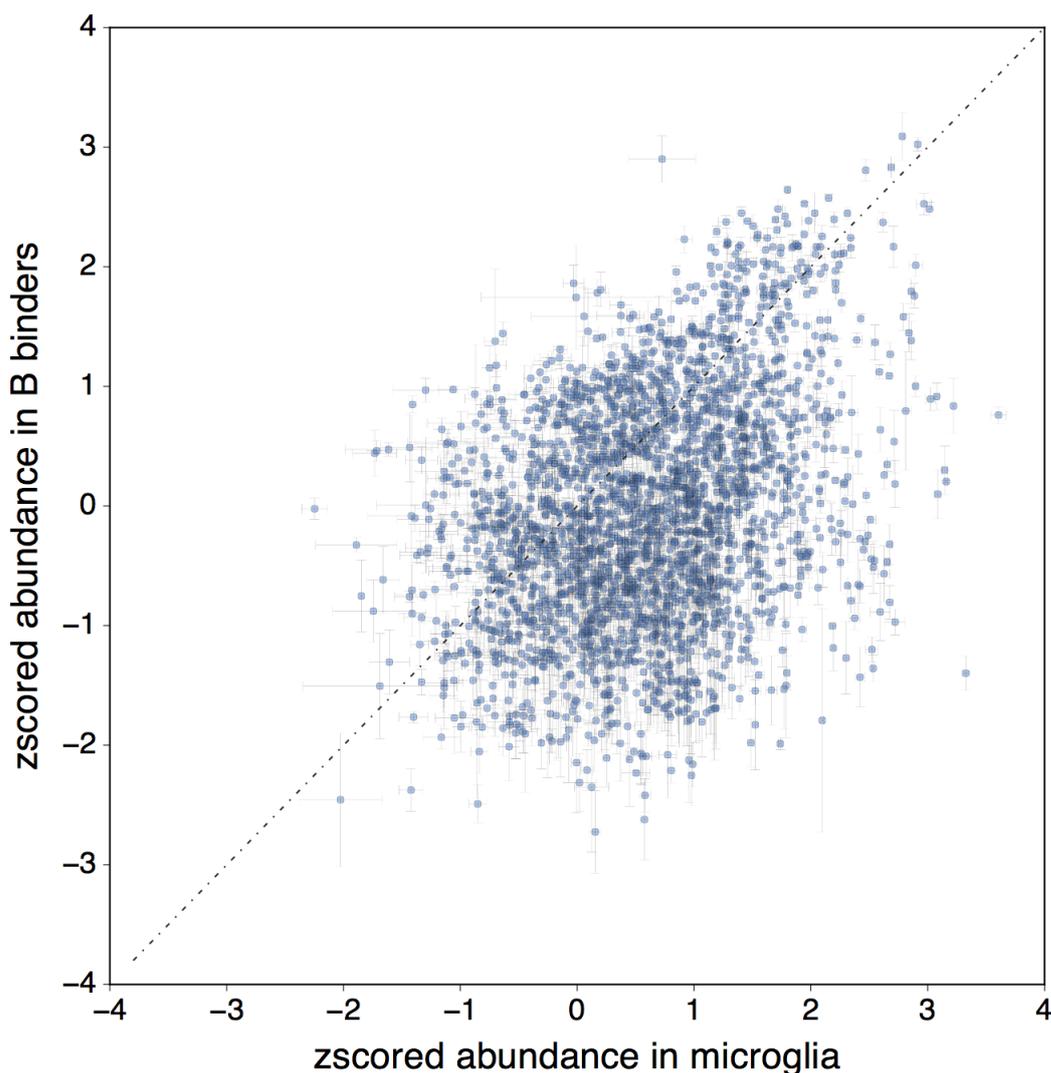


Fig. 4.7

Difference of natural abundance of proteins in microglia is not sufficient to explain the abundance percentages in the interactome

Scatterplot of zscored iBAQ abundances of B binders versus proteins in microglia taken from [207]. Errors are obtained from MAD (median absolute deviation) with normalisation for the zscore transformation.

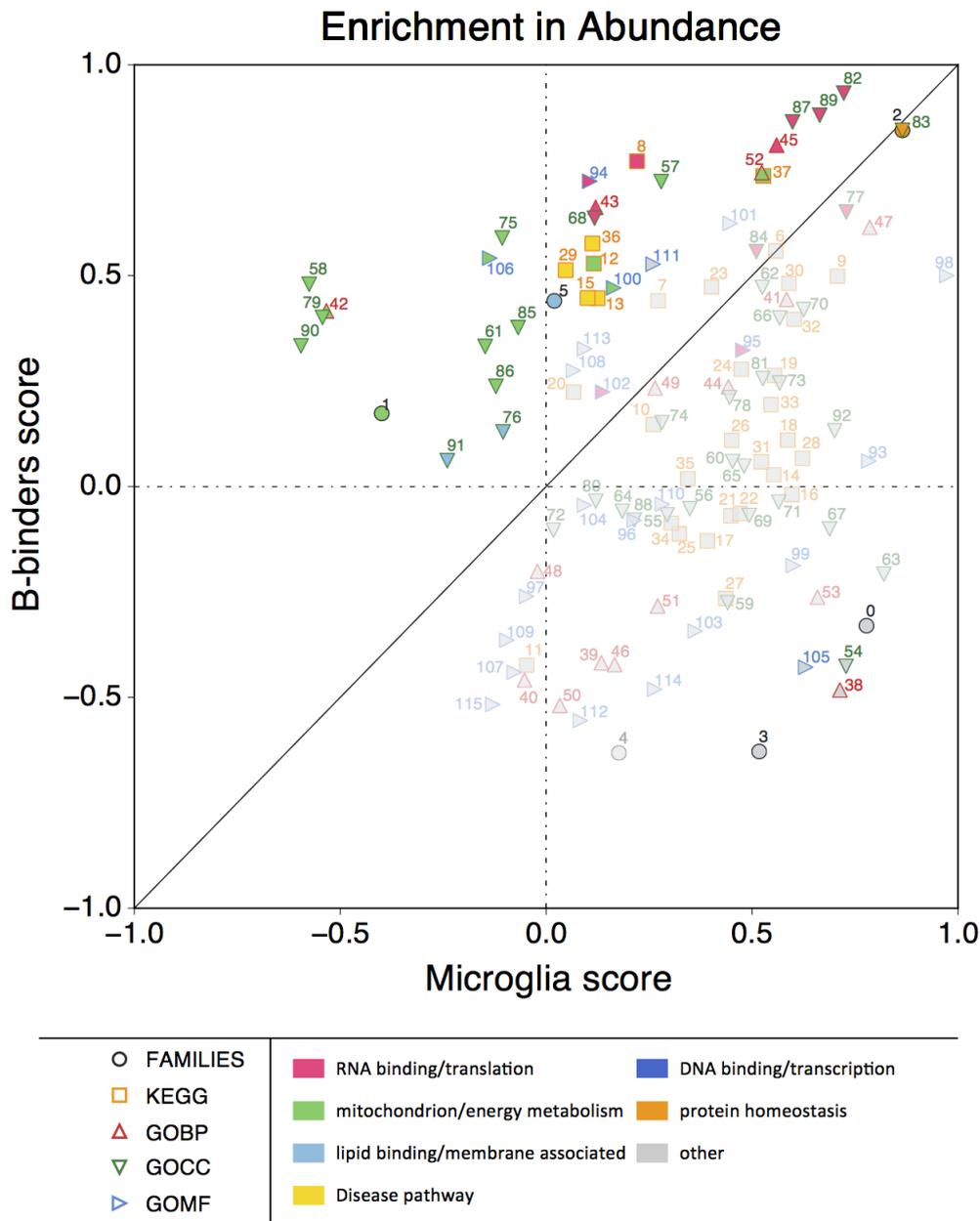
Hence, we asked if there are any biological categories that can significantly differentiate the degree of binding of their components to the oligomers with respect to their natural abundance. Using the mathematical method of 2D functional annotation enrichment developed by Cox and Mann [169] (see Materials and Methods, section 4.2.5), we were able to quantitatively identify which biological properties are differentially expressed in the oligomers sample with respect to the microglia

proteome sample. We tested the 2D enrichment to all functional annotation terms of protein families (1553 entries), KEGG pathways (287 entries), GO biological processes (5192 entries), GO cellular components (950) and GO molecular functions (1809 entries) that were present in the interactome. To calculate the enrichment score in each sample, we used the rankings of abundances of the proteins in the interactome set, and the corresponding ranking of abundances in the microglia set. Only terms that resulted significant ($p < 0.05$ after multiple hypothesis correction with the BH method) from the MANOVA test on the paired 2D enrichment scores were kept. This resulted into 6 protein families, 32 KEGG pathways and 78 GO categories, of which: 16 GO biological processes (GOBP), 39 GO cellular components (GOCC) and 23 GO molecular functions (GOMF), (see legend of Figure 4.8, containing the list of significant terms found for each group).

These annotation terms represents pathways, processes, families, functions or components whose constituents are systematically more abundant or less abundant with respect to all the other proteins, in either the interactome sample or the microglia sample. In particular, we wanted to characterise which functional terms, if any, define sets of proteins that can be considered the most vulnerable species in the cell for the interaction with misfolded oligomers. Vulnerability could come from two behaviours:

- from group of proteins that are not particularly abundant, but have specific characteristic which make them have the strongest affinity to bind the oligomers, and hence would be preferentially chosen as interactors
- from group of proteins which are abundant, so even if their binding affinity is not as high, they would have many more chances to bind the oligomers

Hence, we are interested in characterising which functional terms include proteins that are highly binding the oligomers and low abundant in the microglia (second quadrant in Figure 4.8) as these proteins group would have potentially the strongest affinity to bind the oligomers, and which functional terms define sets whose components are highly binding the oligomers but are also naturally very abundant in cells

**Fig. 4.8****Misfolded oligomers preferentially bind mitochondrial and ribosomal proteins**

2D functional annotation enrichment from abundance rankings of the proteins in the oligomers interactome (B binders) with respect to cellular occurrences in microglia. Terms of protein families (FAMILIES, circles with black borders), KEGG pathways (KEGG, squares with yellow borders) and GO categories (triangles) are present. GO biological processes (GOBP) are triangles with red borders, GO cellular components (GOCC) are inverted triangles with green borders, while GO molecular functions (GOMF) are sideways triangles with blue borders. Only significant terms in the enrichment are reported in the plot (p -value <0.05 , obtained with MANOVA and only relevant regions are highlighted, see section 4.2.5 for explanation). Legend of the annotation terms is found in the next page, with bold labels corresponding to terms found in the relevant regions. Filling colours of relevant terms are matched with the classification of Figure 4.6.

PROTEIN FAMILIES

- 0: 14-3-3 family
- **1: Short-chain dehydrogenases/reductases (SDR) family**
- **2: TCP-1 chaperonin family**
- 3: DOCK family
- 4: Peptidase C19 family
- **5: Small GTPase superfamily, Rab family**

KEGG PATHWAYS

- 6: Pyruvate metabolism
- 7: Fatty acid metabolism
- **8: Ribosome**
- 9: Antigen processing and presentation
- 10: Protein processing in endoplasmic reticulum
- 11: Ubiquitin mediated proteolysis
- **12: Oxidative phosphorylation**
- **13: Alzheimer's disease**
- 14: Salmonella infection
- **15: Huntington's disease**
- 16: Proteasome
- 17: Regulation of actin cytoskeleton
- 18: Leukocyte transendothelial migration
- 19: Biosynthesis of amino acids
- 20: Metabolic pathways
- 21: Tuberculosis
- 22: Platelet activation
- 23: Fatty acid degradation
- 24: Glycolysis / Gluconeogenesis
- 25: PI3K-Akt signaling pathway
- 26: Endocytosis
- 27: Hippo signaling pathway
- 28: Fc gamma R-mediated phagocytosis
- **29: Non-alcoholic fatty liver disease (NAFLD)**
- 30: Carbon metabolism
- 31: Lysosome
- 32: Legionellosis
- 33: Phagosome
- 34: Epstein-Barr virus infection
- 35: Viral carcinogenesis
- **36: Parkinson's disease**
- **37: Citrate cycle (TCA cycle)**

GO BIOLOGICAL PROCESS (GOBP)

- ▲ 38: leukocyte cell-cell adhesion
- ▲ 39: ubiquitin-dependent protein catabolic process
- ▲ 40: protein phosphorylation
- ▲ 41: toxin transport
- ▲ **42: mitochondrial translation**
- ▲ **43: translation**
- ▲ 44: protein folding
- ▲ **45: cytoplasmic translation**
- ▲ 46: positive regulation of GTPase activity
- ▲ 47: cellular response to interleukin-4
- ▲ 48: transcription, DNA-templated
- ▲ 49: intracellular protein transport
- ▲ 50: protein autophosphorylation
- ▲ 51: proteasome-mediated ubiquitin-dependent protein catabolic process
- ▲ **52: tricarboxylic acid cycle**
- ▲ 53: receptor-mediated endocytosis

GO CELLULAR COMPONENT (GOCC)

- ▼ 54: stress fiber
- ▼ 55: lysosome
- ▼ 56: cytosol
- ▼ **57: mitochondrial proton-transporting ATP synthase complex**
- ▼ **58: mitochondrial small ribosomal subunit**
- ▼ 59: cell cortex
- ▼ 60: cell surface
- ▼ **61: mitochondrion**
- ▼ 62: cell body
- ▼ 63: cortical cytoskeleton
- ▼ 64: plasma membrane
- ▼ 65: extracellular space
- ▼ 66: melanosome
- ▼ 67: brush border
- ▼ **68: ribosome**
- ▼ 69: external side of plasma membrane
- ▼ 70: myelin sheath
- ▼ 71: proteasome complex
- ▼ 72: nucleoplasm
- ▼ 73: focal adhesion
- ▼ 74: membrane
- ▼ **75: mitochondrial respiratory chain complex I**
- ▼ **76: endoplasmic reticulum membrane**
- ▼ 77: cytoplasmic ribonucleoprotein granule
- ▼ 78: vesicle
- ▼ **79: mitochondrial large ribosomal subunit**
- ▼ 80: nucleus
- ▼ 81: extracellular exosome
- ▼ **82: small ribosomal subunit**
- ▼ **83: chaperonin-containing T-complex**
- ▼ 84: ribonucleoprotein complex
- ▼ **85: mitochondrial inner membrane**
- ▼ **86: mitochondrial matrix**
- ▼ **87: cytosolic large ribosomal subunit**
- ▼ 88: cytoplasm
- ▼ **89: cytosolic small ribosomal subunit**
- ▼ **90: mitochondrial ribosome**
- ▼ **91: integral component of membrane**
- ▼ 92: proteasome accessory complex

GO MOLECULAR FUNCTION (GOMF)

- ▶ 93: structural constituent of cytoskeleton
- ▶ **94: structural constituent of ribosome**
- ▶ 95: double-stranded RNA binding
- ▶ 96: identical protein binding
- ▶ 97: zinc ion binding
- ▶ 98: MHC class II protein complex binding
- ▶ 99: Rac GTPase binding
- ▶ **100: electron carrier activity**
- ▶ 101: NAD binding
- ▶ 102: poly(A) RNA binding
- ▶ 103: cysteine-type endopeptidase activity
- ▶ 104: ATP binding
- ▶ 105: actin filament binding
- ▶ **106: NADH dehydrogenase (ubiquinone) activity**
- ▶ 107: ligase activity
- ▶ 108: GTP binding
- ▶ 109: chromatin binding
- ▶ 110: protein kinase binding
- ▶ 111: GDP binding
- ▶ 112: ubiquitin-specific protease activity
- ▶ 113: GTPase activity
- ▶ 114: GTPase activator activity
- ▶ 115: protein serine/threonine kinase activity

(values over the bisector in the first quadrant of Figure 4.8). Annotation terms in the area of the second quadrant and upper triangle of the first quadrant in Figure 4.8 define ensemble of proteins that are most vulnerable to the oligomers. We found that overall, RNA-binding proteins (especially involved in translation), protein of the mitochondrion or involved in energy metabolism, proteins associated with neurodegenerative pathways and one family of proteins belonging to the protein homeostasis category are the only categories most vulnerable to the oligomers, compared to all the categories found overrepresented in the sample (Figure 4.6).

In particular, we find that mitochondrial proteins, in particular mitochondrial ribosomal components (numbers 42, 58, 79 and 90 in Figure 4.8) and the respiratory complex I (numbers 75 and 106 in Figure 4.8, 30 protein components) are preferential binders of oligomers, since they are highly abundant in our sample but highly depleted in the microglia one. Indeed, not only we found so many significant annotation terms for mitochondrion in the 2D enrichment, but nearly all these terms are found in the second quadrant, so they are consistently over-abundant in the oligomers with respect to their cellular abundance. One protein family is also found in this area of the 2D enrichment plot, with a low, negative enrichment score in microglia but a positive enrichment in the oligomers sample, which makes it another strong affinity potential binder: it is the short chain dehydrogenases/reductases (SDR) family, also involved in energy metabolism (number 1 in the second quadrant of Figure 4.8), a family of enzymes [265–267] of which we found 24 proteins members in the oligomers interactome. Notably, we do not find any term belonging to this class in the opposite region, i.e. where proteins are abundant in the cell but do not bind strongly to the oligomers (4th quadrant). In this region we find mainly annotation terms related to proteins involved in the structure of the cell, e.g. leukocyte cell-cell adhesion, actin filaments binding, stress fiber and the DOCK family (number 38, 105, 54 and 3 respectively, see Figure 4.8).

Beside the mitochondrion (number 61 in Figure 4.8), another cellular component found depleted in the microglia but upregulated in the oligomers is the membrane: integral components of membrane (number 91), mitochondrial membrane proteins

(number 85 and 86) and proteins of the ER membrane (number 76) are all the remaining annotation terms of the second quadrant.

The only two mitochondrial terms not found in the second quadrant are the KEGG oxidative phosphorylation pathway and the proton-transporting ATP synthase complex (number 12 and 57 respectively in the first quadrant of Figure 4.8), which are extremely overexpressed in the oligomers sample and with medium-high abundance values in microglia respectively. In this region, where only the oligomers binders are highly enriched (see Material and Methods, section 4.2.5) we find 6 of the 7 KEGG pathways (squares in Figure 4.8) that resulted as vulnerable to the oligomers, one of which is the oxidative phosphorylation pathway. Interestingly, four of these pathways are disease-related pathways (yellow squares in the figure), three of which are neurodegenerative: Parkinson's, Alzheimer's and Huntington, number 36, 13 and 15 respectively. The remaining disease pathway is the non-alcoholic fatty liver disease (NAFLD) (number 29 in the figure), which is also connected with oxidative phosphorylation. Also, we find another of the three protein families vulnerable to the oligomers: the Rab family of small GTPases (number 5, light blue circle in the figure).

The last protein family found to be highly enriched in the oligomers is also extremely abundant in the microglia: it is the TCP-1 chaperonin family (8 members found in the sample out of the 8 present in mouse, number 2 in Figure 4.8), involved in the folding and chaperoning of some cytosolic proteins such as actin and tubulin [268–270]. Together with ribosomal components or protein involved with translation (RNA binding proteins, pink filled terms in the first quadrant of Figure 4.8, numbers 43, 68, 94, 8, 45, 87, 89 and 82) and the citrate cycle (TCA, energy metabolism, number 52 and 37 in Figure 4.8) these classes represent the strongest binders to the oligomers, probably mostly due to their high cellular abundance.

Notably, ribosomal components and RNA-binding proteins, while being known to be impaired in Huntington's pathogenesis and other neurodegenerative pathologies [271–273], have also been recently reported to interact with oligomers of Huntingtin

[258]. Our finding from the HypF-N systems hence suggest the presence of a general predisposition of these kind of proteins to bind oligomeric species, and potentially imply another general mechanism of toxicity induced by misfolded oligomers. This is also the case of molecular chaperones, which have been both detected to co-aggregate with aggregated species *in vivo* [90] and to modulate the process of protein aggregation itself [87, 274–280]. Most interestingly, we find evidence of high binding affinity with mitochondrial proteins, from the ribosomal components to the respiratory complex I in the oxidative phosphorylation pathway. Even if mitochondrial impairment and dysfunction has been associated with neurodegeneration, especially in terms of oxidative stress [281–288], the potential interaction of mitochondrial proteins with oligomers was previously unknown. The possible binding to these varied key cellular classes suggest the idea of a multi-hit model of neurodegeneration, in which toxicity might be the result of the gain of function of these vulnerable, key proteins from the direct interaction with the oligomers, or might arise from the loss of function of these protein homeostasis/energy related components upon interaction with the oligomeric species. In both scenarios, this aberrant interaction with multiple protein components could lead to a direct impairment of the protein homeostasis machinery, and a subsequent homeostasis collapse.

4.3 Conclusions

In this chapter, we have characterised the interaction of two structurally different soluble oligomeric species with the protein component of the cell. We have found that the two types of oligomers bind a common pool of proteins, while their structural difference only result into a different binding power to the protein molecules. Type B oligomers, with their hydrophobic patches buried and structured in the assembly, bind much stronger proteins than do type A oligomers, which have the hydrophobic patches unstructured and exposed to the solvent.

More than 2000 proteins are pulled down with the oligomers, defining the shared oligomers interactome. Only one protein resulted as a specific binder to type B

oligomers, while no proteins resulted specific to type A. Of all the biological categories found among the set of oligomers interactors, the most outnumbered ones belong to the classes of RNA-binding and ribosomal proteins, mitochondrial proteins and proteins involved in the energy metabolism or disease pathways, lipid-binding and DNA binding proteins and protein homeostasis components. However, in terms of a quantitative preferential binding and vulnerability of proteins to oligomers, by comparing the natural abundances of proteins in microglia with the abundances of proteins in the oligomers sample, we have found that among the proteins found to interact with the oligomers, we detect a significantly high enrichment and affinity to mitochondrial proteins, ribosomal/RNA-binding components, the TCP-1 chaperonin family and proteins involved in neurodegenerative pathways. The potential aberrant interaction of these key proteins with the oligomers suggest a general mechanism of toxicity induced by the interaction of misfolded oligomers with these protein components of the cell. This mechanism could hold true for the ribosomal components, since interaction with other oligomeric species with ribosomal components has been reported previously this year and impairment of ribosomal biogenesis had been previously associated with neurodegeneration. We do not have evidence yet for a toxicity induced by interaction with mitochondrial proteins, but many studies highlight mitochondrial dysfunction during neurodegeneration.

Taken together, this work provide powerful insights into the relationship between proteins and misfolded oligomers. Also, the information contained in the interactome could represent a valuable reference tool for other kind of studies focused on the interplay between oligomeric species and specific pathways of biological components.

4.4 Materials and Methods

4.4.1 Sample preparation and MS run

The sample preparation involves the steps of the processing of the biological sample until the proteins that bound the oligomers are loaded into the gel, and was carried out by Dr. Benedetta Mannini in the Vendruscolo group. Steps of the process are briefly described in the subsections below.

Preparation of HypF-N oligomers

Protein expression and purification were carried out as described previously [150]. The content of endotoxins in HypF-N protein solution samples was determined by toxin sensor Limulus Amebocyte Lysate (LAL) assay kit (Genscript, Piscataway, NJ, USA) and resulted to be $\sim 0.02EU/ml$. Oligomeric aggregates of HypF-N were prepared by incubating the protein for 4 h at 25 °C and at a concentration of $48\mu M$ in two different experimental conditions:

1. $50mM$ acetate buffer, 12% (v/v) TFE, $2mM$ DTT, pH 5.5 (condition A)
2. $20mM$ TFA, $330mM$ NaCl, pH 1.7 (condition B)

The oligomers were centrifuged at $16100g$ for 10 min and resuspended in buffer.

Cell cultures

Murine N13 microglia cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) F12 supplemented with 10% fetal bovine serum (FBS), 1.0% non-essential amino acids, glutamine and antibiotics. The cell culture was maintained in a 5.0% CO₂ humidified atmosphere at 37 °C and grown until 80% confluence for a maximum of 20 passages.

Pull down assay

Proteins from N13 cell cultures were extracted using the ProteoExtract Native Membrane Protein Extraction Kit (Calbiochem, Darmstadt, Germany) according to the manufacturer's protocol. $10000000\text{cell}/\text{ml}$ of microglial cell lysate and $24\mu\text{M}$ type A or type B oligomers were incubated in isolation or in combination in a final volume of $400\mu\text{L}$ for 1 h at $37\text{ }^\circ\text{C}$ under gentle shaking and then centrifuged at $16100g$ for 10 min. Aliquots of the pellet fractions were subjected to SDS-PAGE using 4-12% Bis-Tris polyacrylamide gels.

MS experiment

The MS experiments were carried out at the Cambridge Centre for Proteomics, which provided us the final mass spectrometry raw files. 1D gel bands were excised and transferred into a 96-well PCR plate. The gel bands were cut into 1mm^2 pieces (10 fractions), destained, reduced (DTT) and alkylated (iodoacetamide) and subjected to enzymatic digestion with trypsin overnight at $37\text{ }^\circ\text{C}$. After digestion, the supernatant was pipetted into a sample vial and loaded onto an autosampler for automated LC-MS/MS analysis.

All LC-MS/MS experiments were performed using a nanoAcquity UPLC (Waters Corp., Milford, MA) system and an LTQ Orbitrap Velos hybrid ion trap mass spectrometer (Thermo Scientific, Waltham, MA). Separation of peptides was performed by reverse-phase chromatography using a Waters reverse-phase nano column (BEH C18, $75\mu\text{m}$ i.d. x 250mm , $1.7\mu\text{m}$ particle size) at flow rate of $300\text{nL}/\text{min}$. Peptides were initially loaded onto a pre-column (Waters UPLC Trap Symmetry C18, $180\mu\text{m}$ i.d x 20mm , $5\mu\text{m}$ particle size) from the nanoAcquity sample manager with 0.1% formic acid for 3 minutes at a flow rate of $5\mu\text{L}/\text{min}$. After this period, the column valve was switched to allow the elution of peptides from the pre-column onto the analytical column. Solvent A was water + 0.1% formic acid and solvent B was acetonitrile + 0.1% formic acid. The linear gradient employed was 3-30% B in 40 minutes (the total run time, including wash/equilibration steps was 60 minutes). The

LC eluant was sprayed into the mass spectrometer by means of a nanospray ion source.

All m/z values of eluting ions were measured in the Orbitrap Velos mass analyzer, set at a resolution of 30000 and scanned from m/z 380-1500. Data dependent scans (Top 20) were employed to automatically isolate and generate fragment ions by collision-induced dissociation (NCE:30%) in the linear ion trap, resulting in the generation of MS/MS spectra. Ions with charge states of 2+ and above were selected for fragmentation.

4.4.2 MS data processing and analysis

MaxQuant processing

Mass spectrometry raw files were processed with the MaxQuant software [205] (version 1.5.3.17). As a protein reference sequence database, we used the swissprot mouse reference proteome (only reviewed entries) obtained on 09 June 2015 from the Uniprot database [211, 251, 212]. Peak lists search were performed automatically with the Andromeda search engine [204] within the MaxQuant environment, which uses a reversed-decoy version of the specified FASTA database to adjust the false discovery rates (FDR) of peptides and proteins. Peptides and proteins false discovery rates (FDR) were kept at default values of 0.01.

The iBAQ quantification was switched on to allow quantification of protein abundances with the iBAQ (intensity-based absolute quantification) method [203]. Only "unique" peptides were used for the quantification, i.e. every peptide detected was used only in one protein group for quantification. This ensures, to a good approximation, independence of the resulting protein abundances, while the independence assumption would not hold if we used non-unique peptides present in several protein groups.

The 10 fractions in each of the 4 biological replicas in each condition (i.e. in the sample with type B or type A oligomers) were processed altogether in the MaxQuant

environment. The feature "Match between runs" was set to True in order to maximise the chances of identification of peptides in case of a low signal, which is the case for the proteins treated with type A oligomers.

The iBAQ method for measuring protein abundance

We define as protein abundance in the oligomers sample of a given biological replica the iBAQ value calculated by MaxQuant for the protein. In the iBAQ quantification, the abundance signal of a given protein is the sum of all the intensities of its detected peptides, divided by the number of theoretically expected peptides for the protein. This normalisation is required for fixing the dependency of the resulting intensity of the protein signal towards its size, as bigger proteins will result in higher abundances. The number of theoretically expected peptides is automatically calculated by MaxQuant with an *in silico* digestion of the protein with the protease used in the experiment.

Proteins in the cell span a wide range of abundances, from few copies to millions copies per cell. This wide distribution of abundances seems to follow a power law, and it is maintained with the intensity-based iBAQ quantification. Hence, when dealing with absolute iBAQ abundances or variables derived from iBAQ values and representing abundances of proteins (abundance percentage, see next section), given the log-normality of the distribution, we apply the logarithmic scale to the data (usually log₁₀) for a better visualisation of the scatterplots (see Figure 4.3, Figure 4.4A and Figure 4.5).

Evaluation of binding differences

The volcano plot in panel B of Figure 4.4B shows the difference in binding that the two types of oligomers exert on proteins (x axis) and its corresponding p-value, calculated with a dependent t-test. The difference in binding, per protein, was calculated as follows. iBAQ abundances were obtained from four biological replicas in the two conditions (binding to type B and type A oligomers). In a given biological

replica, the same batch of proteins from the microglia cells were subjected to interact with type B (condition B) or type A (condition A) oligomers. Hence, we calculated, per biological replica, the pairwise ratio r_i^k of the iBAQ abundances in condition B ($iBAQ_B$) over condition A ($iBAQ_A$) for the replica k of the protein i

$$r_i^k = \frac{iBAQ_B|_i^k}{iBAQ_A|_i^k} \quad (4.1)$$

which give us the binding difference between B and A in a specific biological replica for a given protein. To obtain the difference in binding strength that the two types of oligomers exert on a given protein $(B/A)|_i$, we subsequently averaged the ratios r_i^k over the replicas. When dealing with proteomic data, mean values and standard deviations to represent average behaviours of quantities are not recommended, since they are very sensitive to outliers. Hence, to average the ratios we used the median value of the r_i^k and considered as its error the median absolute deviation (MAD), two measures that are more robust to outliers. Finally, we computed the \log_2 on the $(B/A)|_i$ to show in Figure 4.4B the fold difference in binding between type B and type A oligomers $F(B/A)|_i$. The median fold binding difference that oligomers B have with respect to oligomers A, for a given protein i , is therefore:

$$F(B/A)|_i = \log_2 \left(\text{median}_k \left[\frac{iBAQ_B|_i^k}{iBAQ_A|_i^k} \right] \right) \quad (4.2)$$

where positive values indicates a protein is binding type B oligomers stronger than type A, while negative values indicate the opposite behaviour.

Evaluation of percentages in abundance

From the distribution of protein abundances (iBAQ values) of the oligomers binders, we can estimate the average abundance fraction that each protein contributes to with respect to all the others. We define as binders only proteins detected in at least three replicas. For each oligomeric condition (A or B) we calculated, per biological replica k , the percentage of abundance ($\%_A$ and $\%_B$) of a protein i with respect to the sample:

$$\%_{A|i}^k = \frac{iBAQ_{A|i}^k}{\sum_{j=1}^{N_A^k} iBAQ_{A|j}^k} \cdot 100 \quad (4.3)$$

$$\%_{B|i}^k = \frac{iBAQ_{B|i}^k}{\sum_{j=1}^{N_B^k} iBAQ_{B|j}^k} \cdot 100 \quad (4.4)$$

where N_A^k and N_B^k are the number of proteins in the k biological replica in the sample A and B respectively. The average abundance percentage is then obtained, for each protein i , by taking the median of the $\%_{i}^k$ over the 4 k biological replicas. The error on each abundance percentage is the median absolute deviation (MAD). In a log10 scale, the resulting abundance percentages distributions for all the proteins in the two conditions show a gaussian behaviour. The comparison of the average abundance percentages in the two conditions is reported in the scatterplot of Figure 4.5.

4.4.3 Functional annotation enrichment analysis

To complement the MS analysis with relevant biological insights, we used five functional annotation classes :

- Protein families, from the Uniprot database [211, 289, 251]
- KEGG pathways, from the Kyoto Encyclopedia of Gene and Genomes [290–292]
- and the Gene Ontology (GO) annotation [293–295], subdivided into:
 - Biological processes
 - Cellular components
 - Molecular functions

These functional classes contain lists of protein families, pathways, biological processes etc. which are referred to as "annotation terms" and can contain from one

single protein (very specific term, e.g. "mnmg family" protein family, which has a single protein in mouse) to thousands of proteins (very broad term, e.g. "cytoplasm" cellular component, which nearly reaches 6000 protein).

The most common procedure for enrichment analysis of functional annotations from -omics type of data consist in the calculation of the enrichment ratio (ER) of a given annotation term with its significance [296–299]. The ER is based only on the number of proteins found in the sample belonging to the annotation term, with respect to the number of proteins belonging to the term present in a wider reference ensemble and the relative sizes of the sample and the reference ensemble. In a specific example where the annotation term chosen is the GO cellular component "ribosome", the enrichment ratio will tell how much over-numbered ribosomal components are among the proteins in the interactome (compared to the number of all ribosomal proteins in the proteome of a mouse), and what is the probability to obtain such a number of ribosomal components in the sample by chance.

Calculation of the ER with its significance is a valuable approach for understanding how many and who are the biological constituents of the interactome of oligomers, but it does not provide information on the different contributions that the different terms give when it comes to binding to the oligomers. To this regard, a recent method has been suggested (1D and 2D annotation enrichment, [169]), to evaluate statistical enrichment not only in terms of number of components found for a given annotation term in the sample, but in terms of the values found for the components of the annotation term in the sample. While this method was developed to compare transcriptomics with proteomics data, we extended its formalism to our comparison of protein abundances bound to the oligomers with physiological abundances of proteins in microglia.

The abundance-specific enrichment score with its statistics is described in the subsection "2D enrichment analysis for the preferential binders of oligomers". The next section, "Enrichment ratios for the biological constituents of the interactome", de-

scribes the usual procedure for calculating number-specific enrichment ratios. These analyses were implemented and performed in the python language environment.

Both enrichment calculations of protein binders to the oligomers are reported from the type B oligomers sample. Indeed, we have already shown that we can address the binding of proteins to the two oligomeric types as a single interactome since there is no significant difference in the proteins enriched selectively in the two oligomers. Also, the functional terms significantly overrepresented in the ensemble of proteins detected in at least three replicas for both oligomers types are highly conserved and do not increase if we use as a reference set for the interactome of oligomers the ensemble of proteins that bind at least 3 times type B oligomers binders. Therefore, since abundance data are required for 2D enrichment calculations, for the sake of consistency we consider proteins detected in at least 3 experiments in the type B oligomers sample to be representative of the oligomers interactome in all enrichment calculations (both for enrichment ratio and for 2D enrichment, described below).

Enrichment ratios for the biological constituents of the interactome

The enrichment ratio (ER) measures the overrepresentation of a specific annotation term in the sample under study, compared to a reference population. Our proteome of reference, which we used in the input of MaxQuant as a reference database of protein sequences (see Materials and Methods, section 4.4.2), is the swissprot mouse proteome, containing 23619 entries.

For every protein entry (protein ID), both in the interactome sample (protein binders to type B oligomers in at least 3 replicas) and in the reference proteome, we retrieved all the annotation terms containing the protein, for all five annotation classes. The Uniprot search database [251, 289, 300] was used to obtain the annotation terms associated to each protein (whether existent) for both the protein family class and the GO categories. The Bioservices python package [301] was implemented in the analysis to match a protein ID with all its KEGG pathways.

For the interactome sample, this resulted in 1553 protein families, 287 KEGG pathways and 8125 GO terms, of which: 5192 biological processes (GOBP), 1809 molecular functions (GOMF) and 950 cellular components (GOCC), which were all tested for enrichment.

Next, since we are interested in highlighting in Figure 4.6 the main ontology content enriched in our interactome, we mapped the 8125 GO terms to their corresponding SLIM GO terms [302]. SLIM go terms offer a high-level view on the three gene ontologies (GOBP, GOMF and GOCC) by collecting the most important terms for a proteome without going to the details of the specific fine grained terms. Using the Bioservices python package [301], the GO terms of the reference proteome were mapped to 136 SLIM GO categories, of which 67 SLIM GO biological processes, 40 SLIM GO molecular functions and 29 SLIM GO cellular components.

For of all the annotation terms of the 1553 protein families, 287 KEGG pathways and 136 SLIM GO categories, we calculated the enrichment ratio (ER), the relative maximum enrichment (RME) and the significance with a fisher exact test with adjusted p-values from the Benjamini-Hochberg (BH) correction for multiple hypothesis testing [303]. For a given annotation term t , the ER is calculated as the number of observed components $\#obs(t)|_{sample}$, with respect to the number of expected components $\#pred(t)|_{sample}$, namely:

$$ER(t) = \frac{\#obs(t)|_{sample}}{\#pred(t)|_{sample}} = \frac{\#obs(t)|_{sample}}{\frac{\#obs(t)|_{ref}}{tot|_{ref}} \cdot tot|_{sample}} = \frac{f(t) \text{ in sample}}{f(t) \text{ in ref}} \quad (4.5)$$

where $f(t)$ is the frequency of the annotation term, $\#obs(t)|_{ref}$ is the number of components of t in the reference, and $tot|_{sample}$ and $tot|_{ref}$ are the sample size and reference size respectively.

Given the finite sizes of the sample ensemble and the reference ensemble for a specific annotation term, the ER is bound to have a maximum value. Hence, it can be very informative to evaluate the relative maximum enrichment (RME), which we

define as the ratio between the calculated ER and the maximum ER (ER_{max}) that could be obtained for the specific annotation term. The RME is a quantity in the $[0,1]$ range, where 0 indicates no enrichment, and 1 indicates that the annotation term has reached its maximum value possible in the sample.

The maximum enrichment ER_{max} can differ among annotation terms depending on one condition. If the size of the annotation term t in the reference, i.e. $\#obs(t)|_{ref}$, is bigger than the sample size $tot|_{sample}$, then the maximum number of components that can be observed for t in the sample are equal to the sample size $tot|_{sample}$, and $ER_{max}(t)$:

$$\begin{aligned} & \text{if } tot|_{sample} < \#obs(t)|_{ref} : \\ & ER_{max}(t) = \frac{1}{\frac{\#obs(t)|_{ref}}{tot|_{ref}}} = \frac{tot|_{ref}}{\#obs(t)|_{ref}} \end{aligned} \quad (4.6)$$

depends only on the ratio of the reference size and the size of the annotation term t . The relative maximum enrichment then becomes:

$$RME(t) = \frac{ER(t)}{ER_{max}(t)} = \frac{\#obs(t)|_{sample}}{\frac{\#obs(t)|_{ref}}{tot|_{ref}} \cdot tot|_{sample}} \cdot \frac{\#obs(t)|_{ref}}{tot|_{ref}} = \frac{\#obs(t)|_{sample}}{tot|_{sample}} \quad (4.7)$$

which is simply the frequency $f(t)$ of the annotation term in the sample.

On the contrary, if the size of the annotation term in the reference is smaller than the sample size, then $ER_{max}(t)$ becomes independent of the number of components in the reference:

$$\begin{aligned}
 & \text{if } tot|_{sample} \geq \#obs(t)|_{ref} : \\
 ER_{max}(t) &= \frac{\frac{\#obs(t)|_{ref}}{tot|_{sample}}}{\frac{\#obs(t)|_{ref}}{tot|_{ref}}} = \frac{tot|_{ref}}{tot|_{sample}} \quad (4.8)
 \end{aligned}$$

and remains the simple ratio of the dimension of the two ensembles (the reference and the sample). The relative maximum enrichment in this case becomes the fraction of the components of t in the sample:

$$RME(t) = \frac{ER(t)}{ER_{max}(t)} = \frac{\frac{\#obs(t)|_{sample}}{\#obs(t)|_{ref}} \cdot \frac{tot|_{sample}}{tot|_{ref}}}{\frac{tot|_{sample}}{tot|_{ref}}} = \frac{\#obs(t)|_{sample}}{\#obs(t)|_{ref}} \quad (4.9)$$

Figure 4.6 reports a barchart of the ERs of the annotation terms that resulted very significant ($p < 0.01$) in the protein families, KEGG pathways and SLIM GO terms. Near the edge inside each bar, the RME is shown for each term, and the significance of the ER is indicated with the star notation (**= $p < 0.01$, ***= $p < 10^{-3}$, ****= $p < 10^{-4}$).

2D enrichment analysis for the preferential binders of oligomers

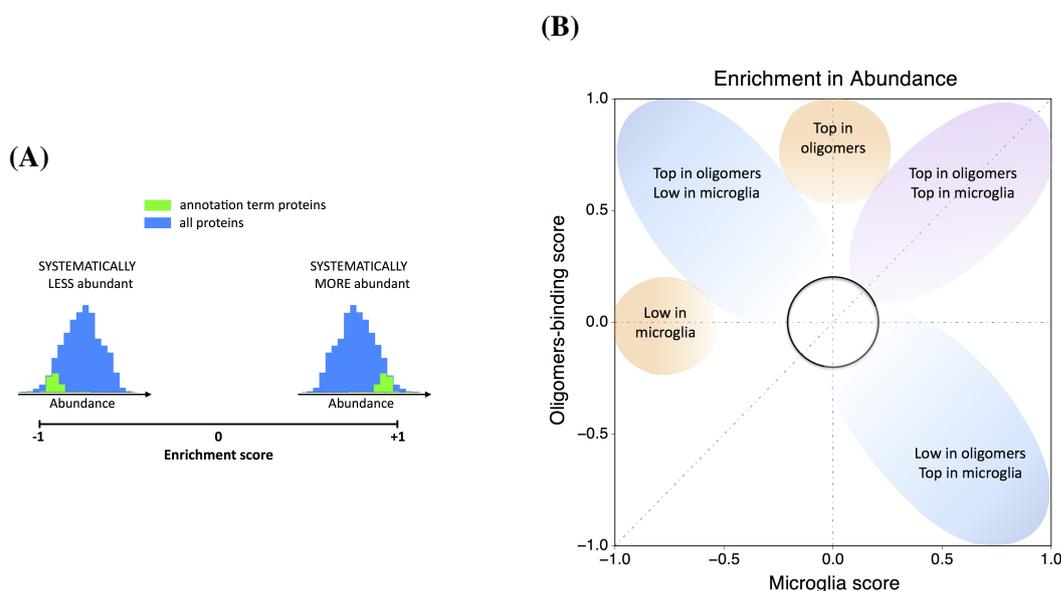
The 2D annotation statistics procedure [169] was adapted to our study and implemented within a python script in the analysis workflow. The 2D annotation enrichment method formulated in [169] is a statistical method that wish to test from two paired distribution whether a subset of entries exhibits a deviated behaviour in at least one of the two distribution. It is a 2-dimensional generalisation applied to the joint distribution of two numerical quantities of the 1D enrichment statistics method, which tests for every subset of a population (in our case an annotation term) whether the corresponding numerical values (in our case protein abundances) tend to be systematically smaller or larger than the global distribution of the population (i.e.

the abundance distribution of all the protein abundances in the sample) (see Figure 4.9A). In this work, protein abundances in microglia and in the interactome were the two numerical quantities for the joint distribution. In particular, we refer to the median percentage in abundances described in section 4.4.2 and section 4.2.5 for both the protein binders sample (type B, at least 3 replicas) and the microglia sample from [207] (mapped to the oligomers, at least 3 replicas), which are plotted after zscore normalisation in Figure 4.7).

The 2D annotation enrichment p-value [169] is defined as the result on the multivariate analysis of variance (MANOVA, e.g. [304]) of the ranked multivariate abundance data, where the data are replaced by ranks in each dimension separately, from 1 to N where the value 1 is given the least abundant protein, and N to the most abundant protein being N the number of proteins in the sample. Another assumption of the formalism, beside the requirement of the ranking transformation for the 2D generalisation non-parametric two-sample test [305], is independence of the values in the distribution [169]. Since both our MS data and the microglia MS data from [207] use unique peptides for protein quantification during the MS processing (see section 4.4.2), the assumption is fulfilled and the method can be applied.

In this work, the 2D annotation enrichment p-value on the ranked abundances of microglia and oligomers binders for a give annotation term t represents the probability for t to be globally over-abundant or depleted in either the microglia sample or the oligomers sample (see Figure 4.9B). While it is also possible to only look for enrichment in large values (one-sided test), we chose to perform a two-sided test and test significance also for deviation to lower abundances since we are also interested in proteins that bind preferentially to the oligomers despite being naturally low abundant.

Since we are now interested in determining the preferential binders of oligomeric species, we considered the full GO categories in the calculations, which comprises broad and fine-grained ontologies. We tested the 1553 protein families, 287 KEGG pathways, 5192 biological processes (GOBP), 1809 molecular functions (GOMF)

**Fig. 4.9****Scheme of the 2D annotation abundance enrichment analysis**

(A) Example scheme of the position-based enriched score which is calculated for each abundance data distribution in the 2D enrichment plot. The enrichment score ranges from -1 to +1 and is calculated for a given annotation term. A value near -1 indicates that the components of the annotation term (green distribution) under study have abundance values that are systematically lower than the remaining values of the distribution. A value close to +1 indicate that the annotation term consists of entries whose abundances are systematically larger than the remaining distribution, while a value of 0 indicates that the term is not distributed differently than the rest of the data. (B) Scheme of the relevant regions of 2D enrichment in our study. The enrichment score for abundance in oligomers binders is plotted against the corresponding enrichment score of abundance in microglia. The empty region near the origin, represented by a void circle, is the result of the cut-off imposed by the significance. The remaining part of the plane rectangle can be subdivided into 8 areas corresponding to correlating, non-correlating and anti-correlating regions, of which we highlighted the 5 most relevant for the purpose of this study. These regions are highlighted to display the general possible behaviour of annotation terms, so shapes and limits should not be taken literally but only as a reference.

and 950 cellular components (GOCC) of the interactome sample. Resulting p -values for the MANOVA of the 2D enrichments were adjusted using the Benjamini-Hochberg method [303] for multiple hypothesis testing in each functional class separately.

For annotation terms that result as significantly deviating the abundance oligomers and microglia level joint distribution, we evaluated a position-enrichment score s which quantifies the abundance deviation in both distributions. Per abundance distribution (either microglia or oligomers), s calculates how much a given annotation term t is ranked higher (or lower) as a group, compared to the ranking of all proteins, hence quantifying the systematical over-abundance or depletion of its components (see Figure 4.9A). It is defined as:

$$s = \frac{2(R_t - R_c)}{n} \quad (4.10)$$

where R_t is the average abundance ranking of the proteins belonging to the annotation term t , R_c is the average ranking of the remaining proteins of the distribution, i.e. the complementary ensemble, and n is the sample size. With such a definition, s ranges from -1, in case of terms present at systematically lower abundances, to +1, for terms made of components with the highest abundances in the sample.

Combining the position-enrichment scores of the microglia distribution and the oligomers distribution, we obtain for each annotation term a number pair (s_x, s_y) , which is the coordinate-wise difference of average abundance ranks in the single distributions. s_x quantifies the enrichment in abundance of the microglia, while s_y is the enrichment in abundance for the oligomers. The resulting 2D enrichments $s = (s_x, s_y)$ are confined to the square $-1 \leq s_x \leq 1$ and $-1 \leq s_y \leq 1$, where the point $(s_x, s_y) = (0, 0)$ corresponds to annotation terms that are not distributed differently from the global distribution of value pairs. Figure 4.9B shows a scheme of the 2D enrichment oligomers/microglia plan with the different regions relevant for our study highlighted in violet, orange and blue. The violet region represents annotation terms that identify sets of proteins that are highly binding the oligomers but also naturally

highly abundant in microglia. The orange areas correspond to terms that are only at the top or at the bottom of the abundance distribution in either one of the two variables. In particular, we are interested in the regions where either proteins are quite strongly binding the oligomers, so they have average values for binding despite being very rare in microglia (left orange circle), or proteins that are just top binders of the oligomers without them being particularly abundant in microglia (top orange circle). Last, the blue regions represent annotation terms showing anti-correlating behaviour: either proteins with the highest affinity to bind the oligomers while being present in low copies naturally, so that they are found extremely enriched in the oligomers sample nonetheless (second quadrant), or proteins that despite being extremely abundant, are not affine to the oligomers and hence are depleted in the interactome (fourth quadrant).

Figure 4.8 in section 4.2.5 shows the results positional enrichment scores of protein abundances in the oligomers sample with respect to protein abundances in microglia, for all the annotation terms that resulted significant.

Chapter 5

Conclusions and Perspectives

In this dissertation I presented a quantitative and high-throughput approach of MS-based proteomics complemented with bioinformatic computational methods. Quantitative proteomics is emerging as a very powerful tool for the direct description at the molecular level of complex biological systems, as it allows *en masse* information on proteins, which perform most of the functions in the cell. The computational methods I used, implemented and contributed to the development of, enable the prediction of the physico-chemical properties of proteins from the knowledge of their amino acid sequences, and characterise the collective proteins behaviour in terms of biological properties like cellular pathways and molecular functions. Combining MS-based experimental data with these biophysical analyses is a strategy that can be used to tackle many biological questions. In this work, I applied this strategy for the study of protein homeostasis and protein aggregation, two major related key concepts in the studying of ageing and neurodegenerative disorders. In order to provide insights into the molecular basis of ageing and neurodegeneration, we rely on the choice of key model systems. The first system is the nematode *C. elegans*, a model organism for ageing research. In chapter 2 and 3 we analysed MS data from nematode *C. elegans* to quantify proteome changes upon ageing and characterise them in terms of physico-chemical principles. We found that ageing in *C. elegans* results in a significant proteome remodelling implicating protein stoichiometry imbalances and widespread

aggregation, which is differently modulated in long-lived and short-lived strains. Although most abundant proteins are less aggregation prone, they contribute the most to the aggregate load, as we found that nearly all the proteins in the adult worm are expressed just above their solubility limits. Despite the wide protein remodelling, we found that the total intracellular protein amount remains constant upon ageing. Nonetheless, aggregate levels are nearly doubling in the same time interval, suggesting that aggregation in the adult nematode is not caused by an increase in abundance, but is a result of a protein-dependent decrease in solubility due to extrinsic factors like disruption of the protein homeostasis network and unbalance in stoichiometries. Also, widespread aggregation occurs in both long-lived and short-lived strains, but proteins in the bigger insoluble deposits of the long-lived strain are more charged, hydrophilic and less aggregation prone than those found in the smaller insoluble deposits of the short-lived strain, suggesting the presence of an extrinsic regulatory mechanism that enhances the aggregation process into forming insoluble inclusion to potentially sequester cytotoxic oligomeric and prefibrillar species. The hypothesis that misfolded oligomers formed at the early stage of the aggregation process may play a key role in the cytotoxicity of neurodegeneration prompted the second part of this work. We performed MS based proteomics to characterise the interaction of proteins with misfolded oligomers. In particular, our model system were two stabilised and well-characterised, structurally different types of oligomers (HypF-N type A and type B). We took advantage of their structural difference to investigate the different biological response upon interaction with proteins. We found that this structural difference modulates the strength of the binding with proteins, but not the specificity. Hence, our model oligomers bind the same pools of proteins, with preferential binding towards mitochondrial and ribosomal proteins, and molecular chaperones.

I believe that the study of the interactome I performed for HypF-N oligomers can bring many advantages. First, it represents a useful collection of information which could be used as a starting point for investigating specific processes mediated by the interaction of oligomers. On this note in particular, it is in my interest to investigate

in the future the interaction of oligomers with mitochondria. To this regard, I am interested in understanding whether the interaction of oligomers with mitochondrial proteins, especially ribosomal components and proteins of the respiratory complex I, can occur *in vivo* and if it can induce cytotoxicity. Also, in order to understand if interaction with proteins intracellularly can be a general mechanism of toxicity for misfolded oligomers, I have planned to study, in collaboration with Dr. Benedetta Mannini, the interactome of other misfolded oligomeric species. Specifically, I am interested in testing the stable oligomers of the proteins mostly involved in neurodegeneration: α -synuclein, A- β 40 and A- β 42 oligomers. I believe that characterising these interactomes will also help us gaining molecular insights into the similarities and differences of neurodegenerative disorders and dementia, and hopefully pointing to a targetable pathway for disease intervention. Lastly, I would like to address not only the interactions these oligomers have with proteins, but also the different response the organism would give to the presence of different misfolded oligomers. To tackle this question, I would like to investigate in collaboration with Michele Perni and Dr. Benedetta Mannini, again with MS based proteomics complemented with our computational methods, the proteome changes in *C. elegans* with time upon exposure to the different oligomers, especially in term of the protein homeostasis network. Also, it would be interesting to check if widespread aggregation can be induced in worms by the presence of the oligomers, and which are the physico-chemical features of the induced aggregates.

From the point of view of the development of computational methods, I would like to develop a method that would take into consideration the uncertainties in the MS data for the functional enrichment analyses, as I believe it could be very important in reducing the presence of potential false positives arising from noisy data. Furthermore, I would like to upgrade the CamSol intrinsic method in order to be able to account for structural correction in the sequence-based aggregation propensity calculations. Such an upgrade would allow the possibility of estimating, indeed, not only the aggregation propensity from the unfolded state, as CamSol intrinsic is giving at the moment, but also the aggregation propensity from the folded state.

In conclusion, I anticipate that proteome-level studies of the type that I have described in this thesis will become increasingly capable of revealing the network of interactions that maintain the proteome in its functional state, and the specific processes that become impaired in ageing and in misfolding diseases. This endeavour will keep me, as well as many other researchers, occupied for several years to come!

References

- [1] L. Partridge and M. Mangel, “Messages from mortality: the evolution of death rates in the old,” *Trends in Ecology & Evolution*, vol. 14, pp. 438–442, Nov. 1999.
- [2] J. H. G. M. van Beek, T. B. L. Kirkwood, and J. B. Bassingthwaite, “Understanding the physiology of the ageing individual: computational modelling of changes in metabolism and endurance,” *Interface Focus*, vol. 6, pp. 20150079–17, Feb. 2016.
- [3] R. J. Colman, T. M. Beasley, J. W. Kemnitz, S. C. Johnson, R. Weindruch, and R. M. Anderson, “Caloric restriction reduces age-related and all-cause mortality in rhesus monkeys,” *Nature Communications*, vol. 5, pp. 1–5, Mar. 2014.
- [4] T. Craig, C. Smelick, R. Tacutu, D. Wuttke, S. H. Wood, H. Stanley, G. Janssens, E. Savitskaya, A. Moskalev, R. Arking, and J. P. de Magalhaes, “The Digital Ageing Atlas: integrating the diversity of age-related changes into a unified resource,” *Nucleic Acids Research*, vol. 43, pp. D873–D878, Jan. 2015.
- [5] H.-m. Chow and K. Herrup, “Genomic integrity and the ageing brain,” *Nature Reviews Neuroscience*, vol. 16, pp. 672–684, Oct. 2015.
- [6] M. R. Klass, “A method for the isolation of longevity mutants in the nematode *Caenorhabditis elegans* and initial results,” *Mechanisms of Ageing and Development*, vol. 22, pp. 279–286, July 1983.
- [7] H. Hsin and C. Kenyon, “Signals from the reproductive system regulate the lifespan of *C. elegans*,” *Nature*, vol. 399, pp. 362–366, May 1999.
- [8] C. J. Kenyon, “The genetics of ageing,” *Nature*, vol. 464, pp. 504–512, Mar. 2010.
- [9] C. López-Otín, M. A. Blasco, L. Partridge, M. Serrano, and G. Kroemer, “The Hallmarks of Aging,” *Cell*, vol. 153, pp. 1194–1217, June 2013.
- [10] J. M. A. Tullet, “DAF-16 target identification in *C. elegans*: past, present and future,” *Biogerontology*, vol. 16, pp. 221–234, Apr. 2015.
- [11] L. R. Lapierre and M. Hansen, “Lessons from *C. elegans*: signaling pathways for longevity,” *Trends in endocrinology and metabolism: TEM*, vol. 23, pp. 591–598, Dec. 2012.

- [12] C. T. Murphy, S. A. McCarroll, C. I. Bargmann, A. Fraser, R. S. Kamath, J. Ahringer, H. Li, and C. Kenyon, "Genes that act downstream of DAF-16 to influence the lifespan of *Caenorhabditis elegans*," *Nature*, vol. 424, pp. 277–283, June 2003.
- [13] H. Ellis, "Genetic control of programmed cell death in the nematode *C. elegans*," *Cell*, vol. 44, pp. 817–829, Mar. 1986.
- [14] M. O. Hengartner, "Genetic control of programmed cell death and aging in the nematode *Caenorhabditis elegans*," *Experimental Gerontology*, vol. 32, pp. 363–374, July 1997.
- [15] "World Population Ageing 2015," *United Nations Department of Economic and Social Affairs Population Division*, Feb. 2016.
- [16] R. L. Nussbaum and C. E. Ellis, "Alzheimer's Disease and Parkinson's Disease," *New England Journal of Medicine*, pp. 1–9, Mar. 2003.
- [17] A. Reeve, E. Simcox, and D. Turnbull, "Ageing and Parkinson's disease: Why is advancing age the biggest risk factor?," *Ageing Research Reviews*, vol. 14, pp. 19–30, Mar. 2014.
- [18] D. A. Bennett, L. A. Beckett, A. M. Murray, K. M. Shannon, C. G. Goetz, D. M. Pilgrim, and D. A. Evans, "Prevalence of Parkinsonian Signs and Associated Mortality in a Community Population of Older People," *New England Journal of Medicine*, vol. 334, no. 2, pp. 71–76, 1996.
- [19] T. J. Collier, N. M. Kanaan, and J. H. Kordower, "Ageing as a primary risk factor for Parkinson's disease: evidence from studies of non-human primates," *Nature Reviews Neuroscience*, vol. 12, pp. 359–366, June 2011.
- [20] P. Scarborough, P. Bhatnagar, K. Wickramasinghe, K. Smolina, C. Mitchell, and M. Rayner, "Coronary heart disease statistics 2010 edition," *British Heart Foundation Statistics Database*, pp. 1–156, Sept. 2010.
- [21] D. M. Lloyd-Jones, M. G. Larson, A. Beiser, and D. Levy, "Lifetime risk of developing coronary heart disease," *Lancet*, vol. 353, no. 9147, pp. 89–92, 1999.
- [22] A. D. Sniderman and C. D. Furberg, "Age as a modifiable risk factor for cardiovascular disease," *Lancet*, vol. 371, no. 9623, pp. 1547–1549, 2008.
- [23] T. Wyss-Coray, "Ageing, neurodegeneration and brain rejuvenation," *Nature*, vol. 539, pp. 180–186, Nov. 2016.
- [24] "2015 Alzheimer's disease facts and figures," *Alzheimer's & dementia : the journal of the Alzheimer's Association*, vol. 12, pp. 459–509, Apr. 2016.
- [25] "2016 Alzheimer's disease facts and figures," *Alzheimer's & dementia : the journal of the Alzheimer's Association*, vol. 12, pp. 459–509, Apr. 2016.
- [26] C. E. Riera and A. Dillin, "Can aging be 'drugged'?" *Nature Medicine*, vol. 21, pp. 1400–1405, Dec. 2015.

- [27] D. M. Morens, J. W. Davis, A. Grandinetti, G. W. Ross, J. S. Popper, and L. R. White, "Epidemiologic observations on Parkinson's disease: Incidence and mortality in a prospective study of middle-aged men," *Neurology*, vol. 46, no. 4, pp. 1044–1050, 1996.
- [28] S. Seshadri, A. Beiser, M. Kelly-Hayes, C. S. Kase, R. Au, W. B. Kannel, and P. A. Wolf, "The lifetime risk of stroke: estimates from the Framingham Study.," *Stroke*, vol. 37, pp. 345–350, Feb. 2006.
- [29] A. Wood-Kaczmar, S. Gandhi, and N. W. Wood, "Understanding the molecular causes of Parkinson's disease.," *Trends in molecular medicine*, vol. 12, pp. 521–528, Nov. 2006.
- [30] T. Niccoli and L. Partridge, "Ageing as a risk factor for disease.," *Current Biology*, vol. 22, pp. R741–52, Sept. 2012.
- [31] L. M. de Lau and M. M. Breteler, "Epidemiology of Parkinson's disease," *The Lancet Neurology*, vol. 5, no. 6, pp. 525–535, 2006.
- [32] B. K. Kennedy, S. L. Berger, A. Brunet, J. Campisi, A. M. Cuervo, E. S. Epel, C. Franceschi, G. J. Lithgow, R. I. Morimoto, J. E. Pessin, T. A. Rando, A. Richardson, E. E. Schadt, T. Wyss-Coray, and F. Sierra, "Geroscience: Linking Aging to Chronic Disease," *Cell*, vol. 159, pp. 709–713, Nov. 2014.
- [33] "World Population Prospects: The 2015 Revision," *United Nations Department of Economic and Social Affairs Population Division*, vol. 36, pp. 854–855, Dec. 2015.
- [34] M. Prince, A. Wimo, M. Guerchet, G.-C. Ali, Y.-T. Wu, and M. Prina, "World Alzheimer Report 2015," tech. rep., Oct. 2015.
- [35] C. M. Dobson, "Alzheimer's disease: addressing a twenty-first century plague," *Rendiconti Lincei*, vol. 26, pp. 251–262, Aug. 2015.
- [36] J. Hardy and S. A. Small, *Neurodegeneration and Dementia*, vol. 1. Chichester, UK: John Wiley & Sons, Ltd, Apr. 2008.
- [37] L. E. Hebert, J. Weuve, P. A. Scherr, and D. A. Evans, "Alzheimer disease in the United States (2010-2050) estimated using the 2010 census," *Neurology*, vol. 80, pp. 1778–1783, May 2013.
- [38] D. J. Selkoe, "Alzheimer's Disease," *Cold Spring Harbor Perspectives in Biology*, vol. 3, pp. a004457–a004457, July 2011.
- [39] M. Prince, A. Comas-Herrera, M. Knapp, M. Guerchet, and M. Karagiannidou, "World Alzheimer Report 2016," *Alzheimer's Disease International*, pp. 1–140, Sept. 2016.
- [40] J. Oeppen and J. W. Vaupel, "Demography. Broken limits to life expectancy.," *Science*, vol. 296, pp. 1029–1031, May 2002.
- [41] A. Wimo and M. Prince, "World Alzheimer Report 2010," *Alzheimer's Disease International*, pp. 1–56, Oct. 2011.

- [42] R. Luengo-Fernandez, J. Leal, and A. M. Gray, "Cost of dementia in the pre-enlargement countries of the European Union.," *Journal of Alzheimer's disease : JAD*, vol. 27, no. 1, pp. 187–196, 2011.
- [43] R. Luengo-Fernandez, J. Leal, and A. M. Gray, "UK research expenditure on dementia, heart disease, stroke and cancer: are levels of spending related to disease burden?," *European journal of neurology*, vol. 19, pp. 149–154, Jan. 2012.
- [44] B. L. Plassman, K. M. Langa, G. G. Fisher, S. G. Heeringa, D. R. Weir, M. B. Ofstedal, J. R. Burke, M. D. Hurd, G. G. Potter, W. L. Rodgers, D. C. Steffens, R. J. Willis, and R. B. Wallace, "Prevalence of dementia in the United States: the aging, demographics, and memory study.," *Neuroepidemiology*, vol. 29, no. 1-2, pp. 125–132, 2007.
- [45] C. P. Hughes, L. Berg, W. L. Danziger, L. A. Coben, and R. L. Martin, "A new clinical scale for the staging of dementia.," *The British journal of psychiatry : the journal of mental science*, vol. 140, pp. 566–572, June 1982.
- [46] J. A. Obeso, M. C. Rodriguez-Oroz, C. G. Goetz, C. Marin, J. H. Kordower, M. Rodriguez, E. C. Hirsch, M. Farrer, A. H. V. Schapira, and G. Halliday, "Missing pieces in the Parkinson's disease puzzle," *Nature Medicine*, vol. 16, pp. 653–661, May 2010.
- [47] C.-W. Zhang, L. Hang, T.-P. Yao, and K.-L. Lim, "Parkin Regulation and Neurodegenerative Disorders.," *Frontiers in aging neuroscience*, vol. 7, p. 248, 2015.
- [48] D. J. Irwin, V. M. Y. Lee, and J. Q. Trojanowski, "Parkinson's disease dementia: convergence of α -synuclein, tau and amyloid- β pathologies.," *Nature Reviews Neuroscience*, vol. 14, pp. 626–636, Sept. 2013.
- [49] S. Przedborski, M. Vila, and V. Jackson-Lewis, "Neurodegeneration: what is it and where are we?," *The Journal of clinical investigation*, vol. 111, pp. 3–10, Jan. 2003.
- [50] I. Alafuzoff, A.-L. Ronnberg, and S. Asikainen-Gustafsson, "Clinical to Histopathological Correlation of the Diagnosis of Dementia," *International Psychogeriatrics*, pp. 1–6, Mar. 1991.
- [51] D. P. Perl, "Neuropathology of Alzheimer's disease.," *The Mount Sinai journal of medicine, New York*, vol. 77, pp. 32–42, Jan. 2010.
- [52] W. W. Seeley, R. K. Crawford, J. Zhou, B. L. Miller, and M. D. Greicius, "Neurodegenerative diseases target large-scale human brain networks.," *Neuron*, vol. 62, pp. 42–52, Apr. 2009.
- [53] R. W. Levenson, V. E. Sturm, and C. M. Haase, "Emotional and behavioral symptoms in neurodegenerative disease: a model for studying the neural bases of psychopathology.," *Annual review of clinical psychology*, vol. 10, pp. 581–606, 2014.

- [54] R. Sulkava, M. Haltia, A. Paetau, J. Wikstrom, and J. Palo, "Accuracy of clinical diagnosis in primary degenerative dementia: correlation with neuropathological findings.," *Journal of Neurology Neurosurgery and Psychiatry*, vol. 46, no. 1, pp. 9–13, 1983.
- [55] D. J. Burn and E. Jaros, "Multiple system atrophy: cellular and molecular pathology.," *Molecular pathology : MP*, vol. 54, pp. 419–426, Dec. 2001.
- [56] R. S. Wilson, E. Segawa, P. A. Boyle, S. E. Anagnos, L. P. Hizel, and D. A. Bennett, "The natural history of cognitive decline in Alzheimer's disease.," *Psychology and Aging*, vol. 27, no. 4, pp. 1008–1017, 2012.
- [57] H. Braak and E. Braak, "Neuropathological staging of Alzheimer-related changes," *Acta Neuropathologica*, vol. 82, pp. 239–259, Sept. 1991.
- [58] D. R. Thal, U. Rub, M. Orantes, and H. Braak, "Phases of Abeta-deposition in the human brain and its relevance for the development of AD," *Neurology*, vol. 58, pp. 1791–1800, June 2002.
- [59] C. M. Dobson, "The Amyloid Phenomenon and Its Links with Human Disease," *Cold Spring Harbor Perspectives in Biology*, pp. a023648–15, Jan. 2017.
- [60] C. M. Dobson, "Protein folding and misfolding," *Nature*, vol. 426, pp. 884–890, Dec. 2003.
- [61] T. P. J. Knowles, M. Vendruscolo, and C. M. Dobson, "The amyloid state and its association with protein misfolding diseases.," *Nature Reviews Molecular Cell Biology*, vol. 15, pp. 384–396, June 2014.
- [62] F. Chiti and C. M. Dobson, "Protein Misfolding, Functional Amyloid, and Human Disease," *Annual review of biochemistry*, vol. 75, pp. 333–366, June 2006.
- [63] D. Eisenberg and M. Jucker, "The Amyloid State of Proteins in Human Diseases," *Cell*, vol. 148, pp. 1188–1203, Mar. 2012.
- [64] E. T. Powers, R. I. Morimoto, A. Dillin, J. W. Kelly, and W. E. Balch, "Biological and Chemical Approaches to Diseases of Proteostasis Deficiency," *Annual review of biochemistry*, vol. 78, pp. 959–991, June 2009.
- [65] P. M. Douglas and A. Dillin, "Protein homeostasis and aging in neurodegeneration," *The Journal of Cell Biology*, vol. 190, pp. 719–729, Sept. 2010.
- [66] O. Sin and E. A. A. Nollen, "Regulation of protein homeostasis in neurodegenerative diseases: the role of coding and non-coding genes," *Cellular and Molecular Life Sciences*, vol. 72, pp. 4027–4047, July 2015.
- [67] M. S. Hipp, S.-H. Park, and F. U. Hartl, "Proteostasis impairment in protein-misfolding and -aggregation diseases.," *Trends in Cell Biology*, vol. 24, pp. 506–514, Sept. 2014.
- [68] S. Kaushik and A. M. Cuervo, "Proteostasis and aging," *Nature Medicine*, vol. 21, pp. 1406–1415, Dec. 2015.

- [69] C. Zhang and A. M. Cuervo, "Restoration of chaperone-mediated autophagy in aging liver improves cellular maintenance and hepatic function.," *Nature Medicine*, vol. 14, pp. 959–965, Sept. 2008.
- [70] F. U. Hartl, "Cellular Homeostasis and Aging," *Annual review of biochemistry*, vol. 85, pp. 1–4, June 2016.
- [71] M. Brehme, C. Voisine, T. Rolland, S. Wachi, J. H. Soper, Y. Zhu, K. Orton, A. Vilella, D. Garza, M. Vidal, H. Ge, and R. I. Morimoto, "A Chaperome Subnetwork Safeguards Proteostasis in Aging and Neurodegenerative Disease," *Cell Reports*, vol. 9, pp. 1135–1150, Nov. 2014.
- [72] W. E. Balch, R. I. Morimoto, A. Dillin, and J. W. Kelly, "Adapting Proteostasis for Disease Intervention," *Science*, vol. 319, pp. 916–919, Feb. 2008.
- [73] R. C. Taylor and A. Dillin, "Aging as an Event of Proteostasis Collapse," *Cold Spring Harbor Perspectives in Biology*, vol. 3, pp. a004440–a004440, May 2011.
- [74] Y. E. Kim, M. S. Hipp, A. Bracher, M. Hayer-Hartl, and F. Ulrich Hartl, "Molecular Chaperone Functions in Protein Folding and Proteostasis," *Annual review of biochemistry*, vol. 82, pp. 323–355, June 2013.
- [75] P. Ciryam, R. I. Morimoto, M. Vendruscolo, C. M. Dobson, and E. P. O'Brien, "In vivo translation rates can substantially delay the cotranslational folding of the Escherichia coli cytosolic proteome.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, pp. E132–40, Jan. 2013.
- [76] P. Ciryam, G. G. Tartaglia, R. I. Morimoto, C. M. Dobson, and M. Vendruscolo, "Widespread Aggregation and Neurodegenerative Diseases Are Associated with Supersaturated Proteins," *Cell Reports*, vol. 5, pp. 781–790, Nov. 2013.
- [77] F. U. Hartl, A. Bracher, and M. Hayer-Hartl, "Molecular chaperones in protein folding and proteostasis," *Nature*, vol. 475, pp. 324–332, July 2011.
- [78] D. Balchin, M. Hayer-Hartl, and F. U. Hartl, "In vivo aspects of protein folding and quality control," *Science*, vol. 353, pp. aac4354–aac4354, June 2016.
- [79] J. Tyedmers, A. Mogk, and B. Bukau, "Cellular strategies for controlling protein aggregation.," *Nature Reviews Molecular Cell Biology*, vol. 11, pp. 777–788, Nov. 2010.
- [80] B. Chen, M. Retzlaff, T. Roos, and J. Frydman, "Cellular Strategies of Protein Quality Control," *Cold Spring Harbor Perspectives in Biology*, vol. 3, pp. a004374–a004374, Aug. 2011.
- [81] V. Liang, M. Ullrich, H. Lam, Y. L. Chew, S. Banister, X. Song, T. Zaw, M. Kassiou, J. Götz, and H. R. Nicholas, "Altered proteostasis in aging and heat shock response in *C. elegans* revealed by analysis of the global and de novo synthesized proteome," *Cellular and Molecular Life Sciences*, vol. 71, pp. 3339–3361, Jan. 2014.

- [82] H. Olzscha, S. M. Schermann, A. C. Woerner, S. Pinkert, M. H. Hecht, G. G. Tartaglia, M. Vendruscolo, M. Hayer-Hartl, F. U. Hartl, and R. M. Vabulas, "Amyloid-like Aggregates Sequester Numerous Metastable Proteins with Essential Cellular Functions," *Cell*, vol. 144, pp. 67–78, Jan. 2011.
- [83] M. Vendruscolo, "Proteome folding and aggregation," *Current Opinion in Structural Biology*, vol. 22, pp. 138–143, Apr. 2012.
- [84] D. L. Stenoien, C. J. Cummings, H. P. Adams, M. G. Mancini, K. Patel, G. N. DeMartino, M. Marcelli, N. L. Weigel, and M. A. Mancini, "Polyglutamine-Expanded Androgen Receptors Form Aggregates That Sequester Heat Shock Proteins, Proteasome Components and SRC-1, and Are Suppressed by the HDJ-2 Chaperone," *Human Molecular Genetics*, vol. 8, pp. 731–741, May 1999.
- [85] N. F. Bence, "Impairment of the Ubiquitin-Proteasome System by Protein Aggregation," *Science*, vol. 292, pp. 1552–1555, May 2001.
- [86] C. I. Holmberg, K. E. Staniszewski, K. N. Mensah, A. Matouschek, and R. I. Morimoto, "Inefficient degradation of truncated polyglutamine proteins by the proteasome," *The EMBO journal*, vol. 23, pp. 4307–4318, Oct. 2004.
- [87] N. R. Jana, "Polyglutamine length-dependent interaction of Hsp40 and Hsp70 family chaperones with truncated N-terminal huntingtin: their role in suppression of aggregation and cellular toxicity," *Human Molecular Genetics*, vol. 9, pp. 2009–2018, Aug. 2000.
- [88] S. Kim, E. A. A. Nollen, K. Kitagawa, V. P. Bindokas, and R. I. Morimoto, "Polyglutamine protein aggregates are dynamic," *Nature Cell Biology*, vol. 4, pp. 826–831, Sept. 2002.
- [89] C. K. Bailey, "Molecular chaperones enhance the degradation of expanded polyglutamine repeat androgen receptor in a cellular model of spinal and bulbar muscular atrophy," *Human Molecular Genetics*, vol. 11, pp. 515–523, Mar. 2002.
- [90] D. M. Walther, P. Kasturi, M. Zheng, S. Pinkert, G. Vecchi, P. Ciryam, R. I. Morimoto, C. M. Dobson, M. Vendruscolo, M. Mann, and F. U. Hartl, "Widespread Proteome Remodeling and Aggregation in Aging *C. elegans*," *Cell*, vol. 161, pp. 919–932, May 2015.
- [91] A. J. Weids, S. Ibstedt, M. J. Tamás, and C. M. Grant, "Distinct stress conditions result in aggregation of proteins with similar properties," *Scientific Reports*, vol. 6, p. 24554, Apr. 2016.
- [92] S. Ibstedt, T. C. Sideri, C. M. Grant, and M. J. Tamas, "Global analysis of protein aggregation in yeast during physiological conditions and arsenite stress," *Biology open*, vol. 3, pp. 913–923, Oct. 2014.
- [93] A. Ben-Zvi, E. A. Miller, and R. I. Morimoto, "Collapse of proteostasis represents an early molecular event in *Caenorhabditis elegans* aging," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, pp. 14914–14919, Sept. 2009.

- [94] D. C. David, N. Ollikainen, J. C. Trinidad, M. P. Cary, A. L. Burlingame, and C. Kenyon, "Widespread Protein Aggregation as an Inherent Part of Aging in *C. elegans*," *PLoS Biology*, vol. 8, Aug. 2010.
- [95] P. Reis-Rodrigues, G. Czerwiec, T. W. Peters, U. S. Evani, S. Alavez, E. A. Gaman, M. Vantipalli, S. D. Mooney, B. W. Gibson, G. J. Lithgow, and R. E. Hughes, "Proteomic analysis of age-dependent changes in protein solubility identifies genes that modulate lifespan," *Aging Cell*, vol. 11, pp. 120–127, Dec. 2011.
- [96] F. Chiti and C. M. Dobson, "Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade.," *Annual review of biochemistry*, vol. 86, pp. 27–68, June 2017.
- [97] J. Hardy, "Genetic Classification of Primary Neurodegenerative Disease," *Science*, vol. 282, pp. 1075–1079, Nov. 1998.
- [98] M. Bucciantini, E. Giannoni, F. Chiti, F. Baroni, L. Formigli, J. Zurdo, N. Taddei, G. Ramponi, C. M. Dobson, and M. Stefani, "Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases," *Nature*, vol. 416, pp. 507–511, Apr. 2002.
- [99] T. J. van Ham, R. Breitling, M. A. Swertz, and E. A. A. Nollen, "Neurodegenerative diseases: Lessons from genome-wide screens in small model organisms.," *EMBO Molecular Medicine*, vol. 1, pp. 360–370, Nov. 2009.
- [100] T. Tyson, J. A. Steiner, and P. Brundin, "Sorting out release, uptake and processing of alpha-synuclein during prion-like spread of pathology.," *Journal of neurochemistry*, Nov. 2015.
- [101] T. Uchihara and B. I. Giasson, "Propagation of alpha-synuclein pathology: hypotheses, discoveries, and yet unresolved questions from experimental and human brain studies," *Acta Neuropathologica*, vol. 131, pp. 49–73, Oct. 2015.
- [102] Y. C. Wong and D. Krainc, " α -synuclein toxicity in neurodegeneration: mechanism and therapeutic strategies," *Nature Medicine*, vol. 23, pp. 1–13, Feb. 2017.
- [103] L. C. Serpell, "Alzheimer's amyloid fibrils: structure and assembly," *Biochimica Et Biophysica Acta*, vol. 1502, pp. 16–30, July 2000.
- [104] K. E. Marshall and L. C. Serpell, "Insights into the Structure of Amyloid Fibrils," *The Open Biology Journal*, vol. 2, no. 2, pp. 185–192, 2010.
- [105] D. E. Otzen, *Amyloid Fibrils and Prefibrillar Aggregates. Molecular and Biological Properties*, John Wiley & Sons, June 2013.
- [106] F. Chiti and C. M. Dobson, "Amyloid formation by globular proteins under native conditions," *Nature Chemical Biology*, vol. 5, pp. 15–22, Jan. 2009.
- [107] M. Sunde, L. C. Serpell, M. Bartlam, P. E. Fraser, M. B. Pepys, and C. C. F. Blake, "Common core structure of amyloid fibrils by synchrotron X-ray diffraction," Edited by F. E. Cohen, *Journal of Molecular Biology*, vol. 273, pp. 729–739, Oct. 1997.

- [108] M. R. Sawaya, S. Sambashivan, R. Nelson, M. I. Ivanova, S. A. Sievers, M. I. Apostol, M. J. Thompson, M. Balbirnie, J. J. W. Wiltzius, H. T. McFarlane, A. Ø. Madsen, C. Riek, and D. Eisenberg, "Atomic structures of amyloid cross-beta spines reveal varied steric zippers.," *Nature*, vol. 447, pp. 453–457, May 2007.
- [109] A. W. P. Fitzpatrick, G. T. Debelouchina, M. J. Bayro, D. K. Clare, M. A. Caporini, V. S. Bajaj, C. P. Jaroniec, L. Wang, V. Ladizhansky, S. A. Müller, C. E. MacPhee, C. A. Waudby, H. R. Mott, A. De Simone, T. P. J. Knowles, H. R. Saibil, M. Vendruscolo, E. V. Orlova, R. G. Griffin, and C. M. Dobson, "Atomic structure and hierarchical assembly of a cross- β amyloid fibril.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, pp. 5468–5473, Apr. 2013.
- [110] C. Wasmer, A. Lange, H. Van Melckebeke, A. B. Siemer, R. Riek, and B. H. Meier, "Amyloid Fibrils of the HET-s(218-289) Prion Form a Solenoid with a Triangular Hydrophobic Core.," *Science*, vol. 319, no. 5869, pp. 1523–1526, 2008.
- [111] T. Lührs, C. Ritter, M. Adrian, D. Riek-Loher, B. Bohrmann, H. Döbeli, D. Schubert, and R. Riek, "3D structure of Alzheimer's amyloid-beta(1-42) fibrils.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, pp. 17342–17347, Nov. 2005.
- [112] R. Tycko, "Solid-state NMR studies of amyloid fibril structure.," *Annual review of physical chemistry*, vol. 62, pp. 279–299, 2011.
- [113] J. L. Jiménez, E. J. Nettleton, M. Bouchard, C. V. Robinson, C. M. Dobson, and H. R. Saibil, "The protofilament structure of insulin amyloid fibrils.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, pp. 9196–9201, July 2002.
- [114] T. P. J. Knowles and M. J. Buehler, "Nanomechanics of functional and pathological amyloid materials.," *Nature nanotechnology*, vol. 6, pp. 469–479, July 2011.
- [115] M. Stefani and C. M. Dobson, "Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution.," *Journal of molecular medicine (Berlin, Germany)*, vol. 81, pp. 678–699, Nov. 2003.
- [116] V. N. Uversky and A. L. Fink, "Conformational constraints for amyloid fibrillation: the importance of being unfolded.," *Biochimica Et Biophysica Acta*, vol. 1698, pp. 131–153, May 2004.
- [117] R. Narayanaswamy, M. Levy, M. Tsechansky, G. M. Stovall, J. D. O'Connell, J. Mirrieles, A. D. Ellington, and E. M. Marcotte, "Widespread reorganization of metabolic enzymes into reversible assemblies upon nutrient starvation.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, pp. 10147–10152, June 2009.
- [118] D. J. Selkoe, "Folding proteins in fatal ways.," *Nature*, vol. 426, pp. 900–904, Dec. 2003.

- [119] D. Thirumalai and G. Reddy, "Protein thermodynamics: Are native proteins metastable?," *Nature chemistry*, vol. 3, pp. 910–911, Dec. 2011.
- [120] A. J. Baldwin, T. P. J. Knowles, G. G. Tartaglia, A. W. Fitzpatrick, G. L. Devlin, S. L. Shammass, C. A. Waudby, M. F. Mossuto, S. Meehan, S. L. Gras, J. Christodoulou, S. J. Anthony-Cahill, P. D. Barker, M. Vendruscolo, and C. M. Dobson, "Metastability of native proteins and the phenomenon of amyloid formation.," *Journal of the American Chemical Society*, vol. 133, pp. 14160–14163, Sept. 2011.
- [121] E. Gazit, "The "Correctly Folded" state of proteins: is it a metastable state?," *Angewandte Chemie (International ed. in English)*, vol. 41, pp. 257–259, Jan. 2002.
- [122] M. Vendruscolo and C. M. Dobson, "Towards complete descriptions of the free-energy landscapes of proteins.," *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, vol. 363, pp. 433–50–discussion 450–2, Feb. 2005.
- [123] M. Vendruscolo and C. M. Dobson, "Structural biology: Protein self-assembly intermediates," *Nature Chemical Biology*, vol. 9, pp. 216–217, Mar. 2013.
- [124] S. Auer, C. M. Dobson, and M. Vendruscolo, "Characterization of the nucleation barriers for protein aggregation and amyloid formation.," *HFSP journal*, vol. 1, pp. 137–146, July 2007.
- [125] S. Auer, M. A. Miller, S. V. Krivov, C. M. Dobson, M. Karplus, and M. Vendruscolo, "Importance of metastable states in the free energy landscapes of polypeptide chains.," *Physical Review Letters*, vol. 99, p. 178104, Oct. 2007.
- [126] S. Auer, C. M. Dobson, M. Vendruscolo, and A. Maritan, "Self-templated nucleation in peptide and protein aggregation.," *Physical Review Letters*, vol. 101, p. 258101, Dec. 2008.
- [127] T. P. J. Knowles, W. Shu, G. L. Devlin, S. Meehan, S. Auer, C. M. Dobson, and M. E. Welland, "Kinetics and thermodynamics of amyloid formation from direct measurements of fluctuations in fibril mass.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, pp. 10016–10021, June 2007.
- [128] F. Chiti, N. Taddei, F. Baroni, C. Capanni, M. Stefani, G. Ramponi, and C. M. Dobson, "Kinetic partitioning of protein folding and aggregation," *Nature Structural Biology*, vol. 9, pp. 137–143, Jan. 2002.
- [129] L. A. Munishkina, E. M. Cooper, V. N. Uversky, and A. L. Fink, "The effect of macromolecular crowding on protein aggregation and amyloid fibril formation.," *Journal of molecular recognition : JMR*, vol. 17, pp. 456–464, Sept. 2004.
- [130] G. G. Tartaglia, S. Pechmann, C. M. Dobson, and M. Vendruscolo, "Life on the edge: a link between gene expression levels and aggregation rates of human proteins," *Trends in Biochemical Sciences*, vol. 32, pp. 204–206, May 2007.

- [131] P. Ciryam, R. Kundra, R. I. Morimoto, C. M. Dobson, and M. Vendruscolo, "Supersaturation is a major driving force for protein aggregation in neurodegenerative diseases," *Trends in Pharmacological Sciences*, vol. 36, pp. 72–77, Feb. 2015.
- [132] A. R. Kinjo and S. Takada, "Competition between Protein Folding and Aggregation with Molecular Chaperones in Crowded Solutions: Insight from Mesoscopic Simulations," *Biophysical Journal*, vol. 85, pp. 3521–3531, Dec. 2003.
- [133] H. Saibil, "Chaperone machines for protein folding, unfolding and disaggregation.," *Nature Reviews Molecular Cell Biology*, vol. 14, pp. 630–642, Oct. 2013.
- [134] T. P. J. Knowles, C. A. Waudby, G. L. Devlin, S. I. A. Cohen, A. Aguzzi, M. Vendruscolo, E. M. Terentjev, M. E. Welland, and C. M. Dobson, "An Analytical Solution to the Kinetics of Breakable Filament Assembly," *Science*, vol. 326, pp. 1533–1537, Dec. 2009.
- [135] S. I. A. Cohen, M. Vendruscolo, M. E. Welland, C. M. Dobson, E. M. Terentjev, and T. P. J. Knowles, "Nucleated polymerization with secondary pathways. I. Time evolution of the principal moments," *The Journal of Chemical Physics*, vol. 135, no. 6, p. 065105, 2011.
- [136] T. P. J. Knowles, A. De Simone, A. W. Fitzpatrick, A. Baldwin, S. Meehan, L. Rajah, M. Vendruscolo, M. E. Welland, C. M. Dobson, and E. M. Terentjev, "Twisting Transition between Crystalline and Fibrillar Phases of Aggregated Peptides," *Physical Review Letters*, vol. 109, p. 158101, Oct. 2012.
- [137] S. I. A. Cohen, M. Vendruscolo, C. M. Dobson, and T. P. J. Knowles, "From Macroscopic Measurements to Microscopic Mechanisms of Protein Aggregation," *Journal of Molecular Biology*, vol. 421, pp. 160–171, Aug. 2012.
- [138] T. Eichner and S. E. Radford, "A diversity of assembly mechanisms of a generic amyloid fold.," *Molecular Cell*, vol. 43, pp. 8–18, July 2011.
- [139] T. R. Jahn and S. E. Radford, "Folding versus aggregation: polypeptide conformations on competing pathways.," *Archives of biochemistry and biophysics*, vol. 469, pp. 100–117, Jan. 2008.
- [140] J. Kaylor, N. Bodner, S. Edridge, G. Yamin, D.-P. Hong, and A. L. Fink, "Characterization of Oligomeric Intermediates in α -Synuclein Fibrillation: FRET Studies of Y125W/Y133F/Y136F α -Synuclein," *Journal of Molecular Biology*, vol. 353, pp. 357–372, Oct. 2005.
- [141] M. M. Apetri, N. C. Maiti, M. G. Zagorski, P. R. Carey, and V. E. Anderson, "Secondary Structure of α -Synuclein Oligomers: Characterization by Raman and Atomic Force Microscopy," *Journal of Molecular Biology*, vol. 355, pp. 63–71, Jan. 2006.
- [142] Y. Yoshimura, Y. Lin, H. Yagi, Y. H. Lee, H. Kitayama, K. Sakurai, M. So, H. Ogi, H. Naiki, and Y. Goto, "Distinguishing crystal-like amyloid fibrils

- and glass-like amorphous aggregates from their kinetics of formation,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 36, pp. 14446–14451, 2012.
- [143] C. M. Dobson, “Principles of protein folding, misfolding and aggregation.,” *Seminars in cell & developmental biology*, vol. 15, pp. 3–16, Feb. 2004.
- [144] E. Karran, M. Mercken, and B. De Strooper, “The amyloid cascade hypothesis for Alzheimer’s disease: an appraisal for the development of therapeutics.,” *Nature Reviews Drug Discovery*, vol. 10, pp. 698–712, Aug. 2011.
- [145] P. T. Lansbury and H. A. Lashuel, “A century-old debate on protein aggregation and neurodegeneration enters the clinic,” *Nature*, vol. 443, pp. 774–779, Oct. 2006.
- [146] C. Haass and D. J. Selkoe, “Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer’s amyloid β -peptide,” *Nature Reviews Molecular Cell Biology*, vol. 8, pp. 101–112, Feb. 2007.
- [147] S. Baglioni, F. Casamenti, M. Bucciantini, L. M. Luheshi, N. Taddei, F. Chiti, C. M. Dobson, and M. Stefani, “Prefibrillar Amyloid Aggregates Could Be Generic Toxins in Higher Organisms,” *Journal of Neuroscience*, vol. 26, pp. 8160–8167, Aug. 2006.
- [148] F. Bemporad and F. Chiti, “Protein Misfolded Oligomers: Experimental Approaches, Mechanism of Formation, and Structure-Toxicity Relationships,” *Chemistry & Biology*, vol. 19, pp. 315–327, Mar. 2012.
- [149] M. Zampagni, R. Cascella, F. Casamenti, C. Grossi, E. Evangelisti, D. Wright, M. Becatti, G. Liguri, B. Mannini, S. Campioni, F. Chiti, and C. Cecchi, “A comparison of the biochemical modifications caused by toxic and non-toxic protein oligomers in cells,” *Journal of Cellular and Molecular Medicine*, vol. 15, pp. 2106–2116, Sept. 2011.
- [150] S. Campioni, B. Mannini, M. Zampagni, A. Pensalfini, C. Parrini, E. Evangelisti, A. Relini, M. Stefani, C. M. Dobson, C. Cecchi, and F. Chiti, “A causative link between the structure of aberrant protein oligomers and their toxicity,” *Nature Chemical Biology*, vol. 6, pp. 140–147, Jan. 2010.
- [151] L. M. Luheshi, G. G. Tartaglia, A.-C. Brorsson, A. P. Pawar, I. E. Watson, F. Chiti, M. Vendruscolo, D. A. Lomas, C. M. Dobson, and D. C. Crowther, “Systematic in vivo analysis of the intrinsic determinants of amyloid Beta pathogenicity.,” *PLoS Biology*, vol. 5, p. e290, Oct. 2007.
- [152] F. Tatini, A. M. Pugliese, C. Traini, S. Niccoli, G. Maraula, T. E. Dami, B. Mannini, T. Scartabelli, F. Pedata, F. Casamenti, and F. Chiti, “Amyloid-beta oligomer synaptotoxicity is mimicked by oligomers of the model protein HypF-N,” *Neurobiology of Aging*, vol. 34, pp. 2100–2109, Sept. 2013.
- [153] G. A. Erikson, D. L. Bodian, M. Rueda, B. Molparia, E. R. Scott, A. A. S.-V. Zeeland, S. E. Topol, N. E. Wineinger, J. E. Niederhuber, E. J. Topol, and A. Torkamani, “Whole-Genome Sequencing of a Healthy Aging Cohort,” *Cell*, vol. 165, pp. 1002–1011, May 2016.

- [154] A. M. Matteini, T. Tanaka, D. Karasik, G. Atzmon, W.-C. Chou, J. D. Eicher, A. D. Johnson, A. M. Arnold, M. L. Callisaya, G. Davies, D. S. Evans, B. Holtfreter, K. Lohman, K. L. Lunetta, M. Mangino, A. V. Smith, J. A. Smith, A. Teumer, L. Yu, D. E. Arking, A. S. Buchman, L. B. Chibinik, P. L. De Jager, D. A. Evans, J. D. Faul, M. E. Garcia, I. Gillham-Nasanya, V. Gudnason, A. Hofman, Y.-H. Hsu, T. Ittermann, L. Lahousse, D. C. Liewald, Y. Liu, L. Lopez, F. Rivadeneira, J. I. Rotter, K. Siggeirsdottir, J. M. Starr, R. Thomson, G. J. Tranah, A. G. Uitterlinden, U. Völker, H. Völzke, D. R. Weir, K. Yaffe, W. Zhao, W. V. Zhuang, J. M. Zmuda, D. A. Bennett, S. R. Cummings, I. J. Deary, L. Ferrucci, T. B. Harris, S. L. R. Kardia, T. Kocher, S. B. Kritchevsky, B. M. Psaty, S. Seshadri, T. D. Spector, V. K. Srikanth, B. G. Windham, M. C. Zillikens, A. B. Newman, J. D. Walston, D. P. Kiel, and J. M. Murabito, “GWAS analysis of handgrip and lower body strength in older adults in the CHARGE consortium,” *Aging Cell*, vol. 15, pp. 792–800, June 2016.
- [155] J. Lund, P. Tedesco, K. Duke, J. Wang, S. K. Kim, and T. E. Johnson, “Transcriptional Profile of Aging in *C. elegans*,” *Current Biology*, vol. 12, pp. 1566–1573, Sept. 2002.
- [156] K. Christensen and M. McGue, “Genetics: Healthy ageing, the genome and the environment.,” *Nature reviews Endocrinology*, vol. 12, pp. 378–380, July 2016.
- [157] A. I. Su, T. Wiltshire, S. Batalov, H. Lapp, K. A. Ching, D. Block, J. Zhang, R. Soden, M. Hayakawa, G. Kreiman, M. P. Cooke, J. R. Walker, and J. B. Hogenesch, “A gene atlas of the mouse and human protein-encoding transcriptomes,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, pp. 6062–6067, Apr. 2004.
- [158] H. Shi, O. Belbin, C. Medway, K. Brown, N. Kalsheker, M. Carrasquillo, P. Proitsi, J. Powell, S. Lovestone, A. Goate, S. Younkin, P. Passmore, Genetic and Environmental Risk for Alzheimer’s Disease Consortium, K. Morgan, and Alzheimer’s Research UK Consortium, “Genetic variants influencing human aging from late-onset Alzheimer’s disease (LOAD) genome-wide association studies (GWAS).,” *Neurobiology of Aging*, vol. 33, pp. 1849.e5–18, Aug. 2012.
- [159] W. R. Jeck, A. P. Siebold, and N. E. Sharpless, “Review: a meta-analysis of GWAS and age-associated diseases.,” *Aging Cell*, vol. 11, pp. 727–731, Oct. 2012.
- [160] P. Ciryam, R. Kundra, R. Freer, R. I. Morimoto, C. M. Dobson, and M. Vendruscolo, “A transcriptional signature of Alzheimer’s disease is associated with a metastable subproteome at risk for aggregation,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, pp. 4753–4758, Apr. 2016.
- [161] R. Freer, P. Sormanni, G. Vecchi, P. Ciryam, C. M. Dobson, and M. Vendruscolo, “A protein homeostasis signature in healthy brains recapitulates tissue vulnerability to Alzheimer’s disease.,” *Science advances*, vol. 2, p. e1600947, Aug. 2016.

- [162] A. A. Podtelezhnikov, K. Q. Tanis, M. Nebozhyn, W. J. Ray, D. J. Stone, and A. P. Loboda, "Molecular insights into the pathogenesis of Alzheimer's disease and its relationship to normal aging.," *Plos One*, vol. 6, no. 12, p. e29610, 2011.
- [163] D. Welter, J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flicek, T. Manolio, L. Hindorff, and H. Parkinson, "The NHGRI GWAS Catalog, a curated resource of SNP-trait associations.," *Nucleic Acids Research*, vol. 42, pp. D1001–6, Jan. 2014.
- [164] R. Tacutu, T. Craig, A. Budovsky, D. Wuttke, G. Lehmann, D. Taranukha, J. Costa, V. E. Fraifeld, and J. P. de Magalhães, "Human Ageing Genomic Resources: integrated databases and tools for the biology and genetics of ageing.," *Nucleic Acids Research*, vol. 41, pp. D1027–33, Jan. 2013.
- [165] P. M. Visscher, M. A. Brown, M. I. McCarthy, and J. Yang, "Five years of GWAS discovery.," *American journal of human genetics*, vol. 90, pp. 7–24, Jan. 2012.
- [166] C. Vogel and E. M. Marcotte, "Insights into the regulation of protein abundance from proteomic and transcriptomic analyses.," *Nature Reviews Genetics*, vol. 13, pp. 227–232, Mar. 2012.
- [167] R. Milo, "What is the total number of protein molecules per cell volume? A call to rethink some published values.," *BioEssays*, vol. 35, pp. 1050–1055, Dec. 2013.
- [168] R. Aebersold and M. Mann, "Mass-spectrometric exploration of proteome structure and function.," *Nature*, vol. 537, pp. 347–355, Sept. 2016.
- [169] J. Cox and M. Mann, "1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data.," *BMC Bioinformatics*, vol. 13 Suppl 16, no. Suppl 16, p. S12, 2012.
- [170] P. Lu, C. Vogel, R. Wang, X. Yao, and E. M. Marcotte, "Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation.," *Nature Biotechnology*, vol. 25, pp. 117–124, Jan. 2007.
- [171] Y.-N. Wei, H.-Y. Hu, G.-C. Xie, N. Fu, Z.-B. Ning, R. Zeng, and P. Khaitovich, "Transcript and protein expression decoupling reveals RNA binding proteins and miRNAs as potential modulators of human aging.," *Genome biology*, vol. 16, p. 41, Feb. 2015.
- [172] M. Bantscheff, S. Lemeer, M. M. Savitski, and B. Kuster, "Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present.," *Analytical and bioanalytical chemistry*, vol. 404, pp. 939–965, Sept. 2012.
- [173] S. Cappadona, P. R. Baker, P. R. Cutillas, A. J. R. Heck, and B. van Breukelen, "Current challenges in software solutions for mass spectrometry-based quantitative proteomics.," *Amino acids*, vol. 43, pp. 1087–1108, Sept. 2012.

- [174] A. Bensimon, A. J. R. Heck, and R. Aebersold, “Mass Spectrometry–Based Proteomics and Network Biology,” *Annual review of biochemistry*, vol. 81, pp. 379–405, July 2012.
- [175] H. Hamzeiy and J. Cox, “What computational non-targeted mass spectrometry-based metabolomics can gain from shotgun proteomics.,” *Current opinion in biotechnology*, vol. 43, pp. 141–146, Feb. 2017.
- [176] P. Mallick and B. Kuster, “Proteomics: a pragmatic perspective.,” *Nature Biotechnology*, vol. 28, pp. 695–709, July 2010.
- [177] L. V. DeSouza and K. W. M. Siu, “Mass spectrometry-based quantification.,” *Clinical biochemistry*, vol. 46, pp. 421–431, Apr. 2013.
- [178] C. C. Wu and J. R. Yates, “The application of mass spectrometry to membrane proteomics,” *Nature Biotechnology*, vol. 21, pp. 262–267, Mar. 2003.
- [179] L. M. F. de Godoy, J. V. Olsen, G. A. de Souza, G. Li, P. Mortensen, and M. Mann, “Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system.,” *Genome biology*, vol. 7, no. 6, p. R50, 2006.
- [180] B. Domon and R. Aebersold, “Review - Mass spectrometry and protein analysis,” *Science*, vol. 312, no. 5771, pp. 212–217, 2006.
- [181] F. Hosp, R. A. Scheltema, H. C. Eberl, N. A. Kulak, E. C. Keilhauer, K. Mayr, and M. Mann, “A Double-Barrel Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS) System to Quantify 96 Interactomes per Day,” *Molecular & Cellular Proteomics*, vol. 14, pp. 2030–2041, July 2015.
- [182] E. C. Keilhauer, M. Y. Hein, and M. Mann, “Accurate Protein Complex Retrieval by Affinity Enrichment Mass Spectrometry (AE-MS) Rather than Affinity Purification Mass Spectrometry (AP-MS),” *Molecular & Cellular Proteomics*, vol. 14, pp. 120–135, Dec. 2014.
- [183] F. Meissner and M. Mann, “Quantitative shotgun proteomics: considerations for a high-quality workflow in immunology.,” *Nature Immunology*, vol. 15, pp. 112–117, Feb. 2014.
- [184] S. B. Azimifar, N. Nagaraj, J. Cox, and M. Mann, “Cell-type-resolved quantitative proteomics of murine liver.,” *Cell metabolism*, vol. 20, pp. 1076–1087, Dec. 2014.
- [185] D. A. Butterfield, L. Gu, F. D. Domenico, and R. A. S. Robinson, “Mass spectrometry and redox proteomics: Applications in disease,” *Mass Spectrometry Reviews*, vol. 33, pp. 277–301, Sept. 2013.
- [186] M. Riffle, G. E. Merrihew, D. Jaschob, V. Sharma, T. N. Davis, W. S. Noble, and M. J. MacCoss, “Visualization and dissemination of multidimensional proteomics data comparing protein abundance during *Caenorhabditis elegans* development.,” *Journal of the American Society for Mass Spectrometry*, vol. 26, pp. 1827–1836, Nov. 2015.

- [187] J. C. Rieckmann, R. Geiger, D. Hornburg, T. Wolf, K. Kveler, D. Jarrossay, F. Sallusto, S. S. Shen-Orr, A. Lanzavecchia, M. Mann, and F. Meissner, "Social network architecture of human immune cells unveiled by quantitative proteomics.," *Nature Immunology*, vol. 2009, p. 439, Mar. 2017.
- [188] J. Koziol, N. Griffin, F. Long, Y. Li, M. Latterich, and J. Schnitzer, "On protein abundance distributions in complex mixtures.," *Proteome science*, vol. 11, no. 1, p. 5, 2013.
- [189] F. F. Gonzalez-Galarza, C. Lawless, S. J. Hubbard, J. Fan, C. Bessant, H. Hermjakob, and A. R. Jones, "A critical appraisal of techniques, software packages, and standards for quantitative proteomic analysis.," *Omics : a journal of integrative biology*, vol. 16, pp. 431–442, Sept. 2012.
- [190] D. M. Walther and M. Mann, "Accurate Quantification of More Than 4000 Mouse Tissue Proteins Reveals Minimal Proteome Changes During Aging," *Molecular & Cellular Proteomics*, vol. 10, pp. M110.004523–M110.004523, Feb. 2011.
- [191] C. C. Wu, M. J. MacCoss, K. E. Howell, and J. R. Yates, "A method for the comprehensive proteomic analysis of membrane proteins," *Nature Biotechnology*, vol. 21, pp. 532–538, Apr. 2003.
- [192] M.-S. Kim, S. M. Pinto, D. Getnet, R. S. Nirujogi, S. S. Manda, R. Chaerkady, A. K. Madugundu, D. S. Kelkar, R. Isserlin, S. Jain, J. K. Thomas, B. Muthusamy, P. Leal-Rojas, P. Kumar, N. A. Sahasrabudde, L. Balakrishnan, J. Advani, B. George, S. Renuse, L. D. N. Selvan, A. H. Patil, V. Nanjappa, A. Radhakrishnan, S. Prasad, T. Subbannayya, R. Raju, M. Kumar, S. K. Sreenivasamurthy, A. Marimuthu, G. J. Sathe, S. Chavan, K. K. Datta, Y. Subbannayya, A. Sahu, S. D. Yelamanchi, S. Jayaram, P. Rajagopalan, J. Sharma, K. R. Murthy, N. Syed, R. Goel, A. A. Khan, S. Ahmad, G. Dey, K. Mudgal, A. Chatterjee, T.-C. Huang, J. Zhong, X. Wu, P. G. Shaw, D. Freed, M. S. Zahari, K. K. Mukherjee, S. Shankar, A. Mahadevan, H. Lam, C. J. Mitchell, S. K. Shankar, P. Satishchandra, J. T. Schroeder, R. Sirdeshmukh, A. Maitra, S. D. Leach, C. G. Drake, M. K. Halushka, T. S. K. Prasad, R. H. Hruban, C. L. Kerr, G. D. Bader, C. A. Iacobuzio-Donahue, H. Gowda, and A. Pandey, "A draft map of the human proteome," *Nature*, vol. 509, pp. 575–581, May 2014.
- [193] Q. Hu, R. J. Noll, H. Li, A. Makarov, M. Hardman, and R. Graham Cooks, "The Orbitrap: a new mass spectrometer.," *Journal of mass spectrometry : JMS*, vol. 40, pp. 430–443, Apr. 2005.
- [194] G. E. Merrihew, C. Davis, B. Ewing, G. Williams, L. Kall, B. E. Frewen, W. S. Noble, P. Green, J. H. Thomas, and M. J. MacCoss, "Use of shotgun proteomics for the identification, confirmation, and correction of *C. elegans* gene annotations," *Genome Research*, vol. 18, pp. 1660–1669, Aug. 2008.
- [195] M. Y. Hein, N. C. Hubner, I. Poser, J. Cox, N. Nagaraj, Y. Toyoda, I. A. Gak, I. Weisswange, J. Mansfeld, F. Buchholz, A. A. Hyman, and M. Mann, "A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances," *Cell*, vol. 163, pp. 712–723, Oct. 2015.

- [196] N. A. Kulak, G. Pichler, I. Paron, N. Nagaraj, and M. Mann, "Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells," *Nature Methods*, vol. 11, pp. 319–324, Feb. 2014.
- [197] M. Mann, "Proteomics for biomedicine: a half-completed journey.," *EMBO Molecular Medicine*, vol. 4, pp. 75–77, Feb. 2012.
- [198] S. E. Ong, "Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics," *Molecular & Cellular Proteomics*, vol. 1, pp. 376–386, May 2002.
- [199] J. R. Wisniewski, M. Y. Hein, J. Cox, and M. Mann, "A "Proteomic Ruler" for Protein Copy Number and Concentration Estimation without Spike-in Standards," *Molecular & Cellular Proteomics*, vol. 13, pp. 3497–3506, Nov. 2014.
- [200] J. Cox and M. Mann, "Quantitative, High-Resolution Proteomics for Data-Driven Systems Biology," *Annual review of biochemistry*, vol. 80, pp. 273–299, July 2011.
- [201] Y. Ishihama, "Exponentially Modified Protein Abundance Index (emPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein," *Molecular & Cellular Proteomics*, vol. 4, pp. 1265–1272, June 2005.
- [202] B. Fabre, T. Lambour, D. Bouyssié, T. Menneteau, B. Monsarrat, O. Burllet-Schiltz, and M.-P. Bousquet-Dubouch, "Comparison of label-free quantification methods for the determination of protein complexes subunits stoichiometry," *EuPA Open Proteomics*, vol. 4, pp. 82–86, Sept. 2014.
- [203] B. Schwanhäusser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach, "Global quantification of mammalian gene expression control," *Nature*, vol. 473, pp. 337–342, May 2011.
- [204] J. Cox, N. Neuhauser, A. Michalski, R. A. Scheltema, J. V. Olsen, and M. Mann, "Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment," *Journal of Proteome Research*, vol. 10, pp. 1794–1805, Apr. 2011.
- [205] J. Cox and M. Mann, "MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification," *Nature Biotechnology*, vol. 26, pp. 1367–1372, Nov. 2008.
- [206] J. Cox, M. Y. Hein, C. A. Luber, I. Paron, N. Nagaraj, and M. Mann, "Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ," *Molecular & Cellular Proteomics*, vol. 13, pp. 2513–2526, Sept. 2014.
- [207] K. Sharma, S. Schmitt, C. G. Bergner, S. Tyanova, N. Kannaiyan, N. Manrique-Hoyos, K. Kongi, L. Cantuti, U.-K. Hanisch, M.-A. Philips, M. J. Rossner, M. Mann, and M. Simons, "Cell type- and brain region-resolved mouse brain proteome.," *Nature neuroscience*, vol. 18, pp. 1819–1831, Dec. 2015.

- [208] A. Christoforou, C. M. Mulvey, L. M. Breckels, A. Geladaki, T. Hurrell, P. C. Hayward, T. Naake, L. Gatto, R. Viner, A. M. Arias, and K. S. Lilley, “A draft map of the mouse pluripotent stem cell spatial proteome,” *Nature Communications*, vol. 7, p. 9992, Jan. 2016.
- [209] P. Sormanni, C. Camilloni, P. Fariselli, and M. Vendruscolo, “The s2D Method: Simultaneous Sequence-Based Prediction of the Statistical Populations of Ordered and Disordered Regions in Proteins,” *Journal of Molecular Biology*, vol. 427, pp. 982–996, Feb. 2015.
- [210] P. Sormanni, F. A. Aprile, and M. Vendruscolo, “The CamSol Method of Rational Design of Protein Mutants with Enhanced Solubility,” *Journal of Molecular Biology*, vol. 427, pp. 478–490, Jan. 2015.
- [211] The UniProt Consortium, “UniProt: the universal protein knowledgebase,” *Nucleic Acids Research*, vol. 45, pp. D158–D169, Jan. 2017.
- [212] R. Leinonen, F. G. Diez, D. Binns, W. Fleischmann, R. Lopez, and R. Apweiler, “UniProt archive,” *Bioinformatics*, vol. 20, pp. 3236–3237, Nov. 2004.
- [213] H. Berman, K. Henrick, and H. Nakamura, “Announcing the worldwide Protein Data Bank,” *Nature Structural Biology*, vol. 10, pp. 980–980, Dec. 2003.
- [214] H. M. Berman, “The Protein Data Bank,” *Nucleic Acids Research*, vol. 28, pp. 235–242, Jan. 2000.
- [215] V. N. Uversky, “A decade and a half of protein intrinsic disorder: Biology still waits for physics,” *Protein science : a publication of the Protein Society*, vol. 22, pp. 693–724, Apr. 2013.
- [216] P. Tompa, “Intrinsically disordered proteins: a 10-year recap,” *Trends in Biochemical Sciences*, vol. 37, pp. 509–516, Dec. 2012.
- [217] V. N. Uversky, “Intrinsically disordered proteins from A to Z,” *The international journal of biochemistry & cell biology*, vol. 43, pp. 1090–1103, Aug. 2011.
- [218] A. K. Dunker, C. J. Brown, J. D. Lawson, L. M. Iakoucheva, and Z. Obradovic, “Intrinsic disorder and protein function,” *Biochemistry*, vol. 41, pp. 6573–6582, May 2002.
- [219] M. M. Babu, “The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease,” *Biochemical Society transactions*, vol. 44, pp. 1185–1200, Oct. 2016.
- [220] S. Chakrabortee, J. S. Byers, S. Jones, D. M. Garcia, B. Bhullar, A. Chang, R. She, L. Lee, B. Fremin, S. Lindquist, and D. F. Jarosz, “Intrinsically Disordered Proteins Drive Emergence and Inheritance of Biological Traits,” *Cell*, vol. 167, pp. 369–381.e12, Oct. 2016.
- [221] P. E. Wright and H. J. Dyson, “Intrinsically disordered proteins in cellular signalling and regulation,” *Nature Reviews Molecular Cell Biology*, vol. 16, pp. 18–29, Jan. 2015.

- [222] R. van der Lee, M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Oldfield, R. V. Pappu, P. Tompa, V. N. Uversky, P. E. Wright, and M. M. Babu, "Classification of intrinsically disordered regions and proteins.," *Chemical Reviews*, vol. 114, pp. 6589–6631, July 2014.
- [223] J. Habchi, P. Tompa, S. Longhi, and V. N. Uversky, "Introducing Protein Intrinsic Disorder," *Chemical Reviews*, vol. 114, pp. 6561–6588, July 2014.
- [224] B. Xue, A. K. Dunker, and V. N. Uversky, "Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life.," *Journal of biomolecular structure & dynamics*, vol. 30, no. 2, pp. 137–149, 2012.
- [225] C. A. Galea, A. A. High, J. C. Obenauer, A. Mishra, C.-G. Park, M. Punta, A. Schlessinger, J. Ma, B. Rost, C. A. Slaughter, and R. W. Kriwacki, "Large-scale analysis of thermostable, mammalian proteins provides insights into the intrinsically disordered proteome.," *Journal of Proteome Research*, vol. 8, pp. 211–226, Jan. 2009.
- [226] M. Sickmeier, J. A. Hamilton, T. LeGall, V. Vacic, M. S. Cortese, A. Tantos, B. Szabo, P. Tompa, J. Chen, V. N. Uversky, Z. Obradovic, and A. K. Dunker, "DisProt: the Database of Disordered Proteins," *Nucleic Acids Research*, vol. 35, pp. D786–D793, Jan. 2007.
- [227] P. Tompa, Z. Dosztanyi, and I. Simon, "Prevalent structural disorder in *E. coli* and *S. cerevisiae* proteomes.," *Journal of Proteome Research*, vol. 5, pp. 1996–2000, Aug. 2006.
- [228] A. K. Dunker, Z. Obradovic, P. Romero, E. C. Garner, and C. J. Brown, "Intrinsic protein disorder in complete genomes.," *Genome informatics. Workshop on Genome Informatics*, vol. 11, pp. 161–171, 2000.
- [229] G. G. Tartaglia and M. Vendruscolo, "The Zyggregator method for predicting protein aggregation propensities.," *Chemical Society reviews*, vol. 37, pp. 1395–1401, July 2008.
- [230] O. Sin, "Aggregation-Promoting Factors in Neurodegenerative Diseases," pp. 1–179, Dec. 2015.
- [231] F. Rousseau, L. Serrano, and J. W. H. Schymkowitz, "How Evolutionary Pressure Against Protein Aggregation Shaped Chaperone Specificity," *Journal of Molecular Biology*, vol. 355, pp. 1037–1047, Feb. 2006.
- [232] E. Monsellier, M. Ramazzotti, N. Taddei, and F. Chiti, "Aggregation propensity of the human proteome.," *PLoS Computational Biology*, vol. 4, p. e1000199, Oct. 2008.
- [233] E. Monsellier, M. Ramazzotti, P. P. de Laureto, G. G. Tartaglia, N. Taddei, A. Fontana, M. Vendruscolo, and F. Chiti, "The Distribution of Residues in a Polypeptide Sequence Is a Determinant of Aggregation Optimized by Evolution," *Biophysical Journal*, vol. 93, pp. 4382–4391, Dec. 2007.

- [234] E. Monsellier and F. Chiti, "Prevention of amyloid-like aggregation as a driving force of protein evolution," *EMBO reports*, vol. 8, pp. 737–742, Aug. 2007.
- [235] J. Reumers, S. Maurer-Stroh, J. Schymkowitz, and F. Rousseau, "Protein sequences encode safeguards against aggregation," *Human mutation*, vol. 30, pp. 431–437, Mar. 2009.
- [236] H.-X. Zhou, G. Rivas, and A. P. Minton, "Macromolecular Crowding and Confinement: Biochemical, Biophysical, and Potential Physiological Consequences *," *Annual Review of Biophysics*, vol. 37, pp. 375–397, June 2008.
- [237] T. Gidalevitz, V. Prahlad, and R. I. Morimoto, "The stress of protein misfolding: from single cells to multicellular organisms.," *Cold Spring Harbor Perspectives in Biology*, vol. 3, June 2011.
- [238] A. Ben-Zvi, E. A. Miller, and R. I. Morimoto, "Collapse of proteostasis represents an early molecular event in *Caenorhabditis elegans* aging," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, pp. 14914–14919, Sept. 2009.
- [239] V. Castillo, R. Graña-Montes, and S. Ventura, "The aggregation properties of *Escherichia coli* proteins associated with their cellular abundance," *Biotechnology Journal*, vol. 6, pp. 752–760, Apr. 2011.
- [240] J. Gsponer and M. M. Babu, "Cellular Strategies for Regulating Functional and Nonfunctional Protein Aggregation," *Cell Reports*, vol. 2, pp. 1425–1437, Nov. 2012.
- [241] A. G. Cashikar, M. Duennwald, and S. L. Lindquist, "A chaperone pathway in protein disaggregation. Hsp26 alters the nature of protein aggregates to facilitate reactivation by Hsp104.," *The Journal of Biological Chemistry*, vol. 280, pp. 23869–23875, June 2005.
- [242] E. Laskowska, A. Wawrzynów, and A. Taylor, "IbpA and IbpB, the new heat-shock proteins, bind to endogenous *Escherichia coli* proteins aggregated intracellularly by heat shock," *Biochimie*, vol. 78, no. 2, pp. 117–122, 1996.
- [243] E. Basha, G. J. Lee, L. A. Breci, A. C. Hausrath, N. R. Buan, K. C. Giese, and E. Vierling, "The Identity of Proteins Associated with a Small Heat Shock Protein during Heat Stress in Vivo Indicates That These Chaperones Protect a Wide Range of Cellular Functions," *The Journal of Biological Chemistry*, vol. 279, pp. 7566–7575, Feb. 2004.
- [244] A. Mogk, E. Deuerling, S. Vorderwülbecke, E. Vierling, and B. Bukau, "Small heat shock proteins, ClpB and the DnaK system form a functional triade in reversing protein aggregation," *Molecular Microbiology*, vol. 50, pp. 585–595, Aug. 2003.
- [245] S. Specht, S. B. M. Miller, A. Mogk, and B. Bukau, "Hsp42 is required for sequestration of protein aggregates into deposition sites in *Saccharomyces cerevisiae*," *The Journal of Cell Biology*, vol. 195, pp. 617–629, Nov. 2011.

- [246] M. Arrasate, S. Mitra, E. S. Schweitzer, M. R. Segal, and S. Finkbeiner, "Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death.," *Nature*, vol. 431, pp. 805–810, Oct. 2004.
- [247] P. M. Douglas, S. Treusch, H.-Y. Ren, R. Halfmann, M. L. Duennwald, S. Lindquist, and D. M. Cyr, "Chaperone-dependent amyloid assembly protects cells from prion toxicity.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, pp. 7206–7211, May 2008.
- [248] A. M. R. de Graff, M. J. Hazoglou, and K. A. Dill, "Highly Charged Proteins: The Achilles' Heel of Aging Proteomes," *Structure/Folding and Design*, vol. 24, pp. 329–336, Feb. 2016.
- [249] T. Fleckenstein, A. Kastenmüller, M. L. Stein, C. Peters, M. Daake, M. Krause, D. Weinfurter, M. Haslbeck, S. Weinkauf, M. Groll, and J. Buchner, "The Chaperone Activity of the Developmental Small Heat Shock Protein Sip1 Is Regulated by pH- Dependent Conformational Changes," *Molecular Cell*, vol. 58, pp. 1067–1078, June 2015.
- [250] M. J. Vos, J. Hageman, S. Carra, and H. H. Kampinga, "Structural and Functional Diversities between Members of the Human HSPB, HSPH, HSPA, and DNAJ Chaperone Families †," *Biochemistry*, vol. 47, pp. 7001–7011, July 2008.
- [251] The UniProt Consortium, "Reorganizing the protein space at the Universal Protein Resource (UniProt)," *Nucleic Acids Research*, vol. 40, pp. D71–D75, Dec. 2011.
- [252] T. N. Petersen, S. Brunak, G. von Heijne, and H. Nielsen, "SignalP 4.0: discriminating signal peptides from transmembrane regions," *Nature Methods*, vol. 8, pp. 785–786, Sept. 2011.
- [253] P. Horton, K. J. Park, T. Obayashi, N. Fujita, H. Harada, C. J. Adams-Collier, and K. Nakai, "WoLF PSORT: protein localization predictor," *Nucleic Acids Research*, vol. 35, pp. W585–W587, May 2007.
- [254] B. Mannini, E. Mulvihill, C. Sgromo, R. Cascella, R. Khodarahmi, M. Ramazzotti, C. M. Dobson, C. Cecchi, and F. Chiti, "Toxicity of Protein Oligomers Is Rationalized by a Function Combining Size and Surface Hydrophobicity," *ACS Chemical Biology*, vol. 9, pp. 2309–2317, Oct. 2014.
- [255] N. P. Reynolds, A. Soragni, M. Rabe, D. Verdes, E. Liverani, S. Handschin, R. Riek, and S. Seeger, "Mechanism of membrane interaction and disruption by α -synuclein.," *Journal of the American Chemical Society*, vol. 133, pp. 19366–19375, Dec. 2011.
- [256] M. Nors Perderson, V. Foderà, I. Horvath, A. van Maarschalkerweerd, K. Nørgaard Toft, C. Weise, F. Almqvist, M. Wolf-Watz, P. Wittung-Stafshede, and B. Vestergaard, "Direct Correlation Between Ligand-Induced α -Synuclein Oligomers and Amyloid-like Fibril Growth.," *Scientific Reports*, vol. 5, p. 10422, May 2015.

- [257] N. B. Last, E. Rhoades, and A. D. Miranker, "Islet amyloid polypeptide demonstrates a persistent capacity to disrupt membrane integrity.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, pp. 9460–9465, June 2011.
- [258] Y. E. Kim, F. Hosp, F. Frottin, H. Ge, M. Mann, M. Hayer-Hartl, and F. U. Hartl, "Soluble Oligomers of PolyQ-Expanded Huntingtin Target a Multiplicity of Key Cellular Factors," *Molecular Cell*, vol. 63, pp. 951–964, Sept. 2016.
- [259] N. B. Last and A. D. Miranker, "Common mechanism unites membrane poration by amyloid and antimicrobial peptides.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, pp. 6382–6387, Apr. 2013.
- [260] H. A. Lashuel, D. Hartley, B. M. Petre, T. Walz, and P. T. Lansbury, "Neurodegenerative disease: amyloid pores from pathogenic mutations.," *Nature*, vol. 418, pp. 291–291, July 2002.
- [261] H. A. Lashuel, B. M. Petre, J. Wall, M. Simon, R. J. Nowak, T. Walz, and P. T. Lansbury, " α -Synuclein, Especially the Parkinson's Disease-associated Mutants, Forms Pore-like Annular and Tubular Protofibrils," *Journal of Molecular Biology*, vol. 322, no. 5, pp. 1089–1102, 2002.
- [262] L. Giehm, D. I. Svergun, D. E. Otzen, and B. Vestergaard, "Low-resolution structure of a vesicle disrupting α -synuclein oligomer that accumulates during fibrillation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, pp. 3246–3251, Feb. 2011.
- [263] P. Cao, A. Abedini, H. Wang, L.-H. Tu, X. Zhang, A. M. Schmidt, and D. P. Raleigh, "Islet amyloid polypeptide toxicity and membrane interactions.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, pp. 19279–19284, Nov. 2013.
- [264] S. Chasseigneaux, C. Clamagirand, L. Huguet, L. Gorisse-Hussonnois, C. Rose, and B. Allinquant, "Cytoplasmic SET induces tau hyperphosphorylation through a decrease of methylated phosphatase 2A.," *BMC neuroscience*, vol. 15, p. 82, June 2014.
- [265] Y. Kallberg, U. Oppermann, and B. Persson, "Classification of the short-chain dehydrogenase/reductase superfamily using hidden Markov models.," *FEBS Journal*, vol. 277, pp. 2375–2386, May 2010.
- [266] Y. Kallberg, U. Oppermann, H. Jörnvall, and B. Persson, "Short-chain dehydrogenases/reductases (SDRs)," *European Journal of Biochemistry*, vol. 269, pp. 4409–4417, Sept. 2002.
- [267] U. Oppermann, C. Filling, M. Hult, N. Shafqat, X. Q. Wu, M. Lindh, J. Shafqat, E. Nordling, Y. Kallberg, B. Persson, and H. Jörnvall, "Short-chain dehydrogenases/reductases (SDR): the 2002 update," *Chemico-Biological Interactions*, vol. 143, pp. 247–253, 2003.

- [268] K. I. Brackley and J. Grantham, "Activities of the chaperonin containing TCP-1 (CCT): implications for cell cycle progression and cytoskeletal organisation.," *Cell stress & chaperones*, vol. 14, pp. 23–31, Jan. 2009.
- [269] R. Melki, G. Batelier, S. Soulié, and R. C. Williams, "Cytoplasmic chaperonin containing TCP-1: structural and functional characterization.," *Biochemistry*, vol. 36, pp. 5817–5826, May 1997.
- [270] H. Kubota, G. Hynes, and K. Willison, "The chaperonin containing t-complex polypeptide 1 (TCP-1). Multisubunit machinery assisting in protein folding and assembly in the eukaryotic cytosol.," *European Journal of Biochemistry*, vol. 230, pp. 3–16, May 1995.
- [271] B. P. Culver, J. N. Savas, S. K. Park, J. H. Choi, S. Zheng, S. O. Zeitlin, J. R. Yates, and N. Tanese, "Proteomic analysis of wild-type and mutant huntingtin-associated proteins in mouse brains identifies unique interactions and involvement in protein synthesis.," *The Journal of Biological Chemistry*, vol. 287, pp. 21599–21614, June 2012.
- [272] A. Hodges, "Regional and cellular gene expression changes in human Huntington's disease brain," *Human Molecular Genetics*, vol. 15, pp. 965–977, Jan. 2006.
- [273] J. Yang, X. Hao, X. Cao, B. Liu, and T. Nystrom, "Spatial sequestration and detoxification of Huntingtin by the ribosome quality control complex," *eLife*, vol. 5, 2016.
- [274] D. Cox, E. Selig, M. D. W. Griffin, J. A. Carver, and H. Ecroyd, "Small Heat-shock Proteins Prevent α -Synuclein Aggregation via Transient Interactions and Their Efficacy Is Affected by the Rate of Aggregation," *The Journal of Biological Chemistry*, vol. 291, pp. 22618–22629, Oct. 2016.
- [275] S. Tam, R. Geller, C. Spiess, and J. Frydman, "The chaperonin TRiC controls polyglutamine aggregation and toxicity through subunit-specific interactions.," *Nature Cell Biology*, vol. 8, pp. 1155–1162, Oct. 2006.
- [276] B. Mannini, R. Casella, M. Zampagni, M. van Waarde-Verhagen, S. Meehan, C. Roodveldt, S. Campioni, M. Boninsegna, A. Penco, A. Relini, H. H. Kampinga, C. M. Dobson, M. R. Wilson, C. Cecchi, and F. Chiti, "Molecular mechanisms used by chaperones to reduce the toxicity of aberrant protein oligomers.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, pp. 12479–12484, July 2012.
- [277] P. Arosio, M. Vendruscolo, C. M. Dobson, and T. P. J. Knowles, "Chemical kinetics for drug discovery to combat protein aggregation diseases.," *Trends in Pharmacological Sciences*, vol. 35, pp. 127–135, Mar. 2014.
- [278] M. A. Wright, F. A. Aprile, P. Arosio, M. Vendruscolo, C. M. Dobson, and T. P. J. Knowles, "Biophysical approaches for the study of interactions between molecular chaperones and protein aggregates.," *Chemical communications*, vol. 51, pp. 14425–14434, Oct. 2015.

- [279] V. Kakkar, C. Månsson, E. P. de Mattos, S. Bergink, M. van der Zwaag, M. A. W. H. van Waarde, N. J. Kloosterhuis, R. Melki, R. T. P. van Cruchten, S. Al-Karadaghi, P. Arosio, C. M. Dobson, T. P. J. Knowles, G. P. Bates, J. M. van Deursen, S. Linse, B. van de Sluis, C. Emanuelsson, and H. H. Kampinga, "The S/T-Rich Motif in the DNAJB6 Chaperone Delays Polyglutamine Aggregation and the Onset of Disease in a Mouse Model," *Molecular Cell*, vol. 62, pp. 272–283, Apr. 2016.
- [280] P. Arosio, T. C. T. Michaels, S. Linse, C. Månsson, C. Emanuelsson, J. Presto, J. Johansson, M. Vendruscolo, C. M. Dobson, and T. P. J. Knowles, "Kinetic analysis reveals the diversity of microscopic mechanisms through which molecular chaperones suppress amyloid formation.," *Nature Communications*, vol. 7, p. 10948, Mar. 2016.
- [281] M. E. Witte, J. J. G. Geurts, H. E. de Vries, P. van der Valk, and J. van Horssen, "Mitochondrial dysfunction: A potential link between neuroinflammation and neurodegeneration?," *Mitochondrion*, vol. 10, pp. 411–418, Aug. 2010.
- [282] M. J. Baker, C. S. Palmer, and D. Stojanovski, "Mitochondrial protein quality control in health and disease.," *British journal of pharmacology*, vol. 171, pp. 1870–1889, Apr. 2014.
- [283] M. H. Yan, X. Wang, and X. Zhu, "Mitochondrial defects and oxidative stress in Alzheimer disease and Parkinson disease," *Free Radical Biology and Medicine*, vol. 62, pp. 90–101, Sept. 2013.
- [284] V. S. Van Laar and S. B. Berman, "Mitochondrial dynamics in Parkinson's disease," *Experimental Neurology*, vol. 218, pp. 247–256, Aug. 2009.
- [285] M. Morán, D. Moreno-Lastres, L. Marín-Buera, J. Arenas, M. A. Martín, and C. Ugalde, "Mitochondrial respiratory chain dysfunction Implications in neurodegeneration," *Free Radical Biology and Medicine*, vol. 53, pp. 595–609, Aug. 2012.
- [286] S. C. Correia, G. Perry, and P. I. Moreira, "Mitochondrial traffic jams in Alzheimer's disease - pinpointing the roadblocks," *Biochimica Et Biophysica Acta*, vol. 1862, pp. 1909–1917, Oct. 2016.
- [287] P. Guedes-Dias, B. R. Pinho, T. R. Soares, J. de Proença, M. R. Duchon, and J. M. A. Oliveira, "Mitochondrial dynamics and quality control in Huntington's disease," *Neurobiology of Disease*, vol. 90, pp. 51–57, June 2016.
- [288] K. F. Winklhofer and C. Haass, "Mitochondrial dysfunction in Parkinson's disease," *Biochimica Et Biophysica Acta*, vol. 1802, pp. 29–44, Jan. 2010.
- [289] S. Pundir, M. Magrane, M. J. Martin, C. O'Donovan, and UniProt Consortium, "Searching and Navigating UniProt Databases.," *Current protocols in bioinformatics*, vol. 50, pp. 1.27.1–10, June 2015.
- [290] M. Kanehisa and S. Goto, "KEGG: Kyoto Encyclopedia of Genes and Genomes," *Nucleic Acids Research*, vol. 28, no. 1, pp. 27–30, 2000.

- [291] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe, "KEGG as a reference resource for gene and protein annotation.," *Nucleic Acids Research*, vol. 44, pp. D457–62, Jan. 2016.
- [292] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima, "KEGG: new perspectives on genomes, pathways, diseases and drugs.," *Nucleic Acids Research*, vol. 45, pp. D353–D361, Jan. 2017.
- [293] R. P. Huntley, T. Sawford, P. Mutowo-Meullenet, A. Shypitsyna, C. Bonilla, M. J. Martin, and C. O'Donovan, "The GOA database: gene Ontology annotation updates for 2015.," *Nucleic Acids Research*, vol. 43, pp. D1057–63, Jan. 2015.
- [294] Gene Ontology Consortium, "Gene Ontology Consortium: going forward.," *Nucleic Acids Research*, vol. 43, pp. D1049–56, Jan. 2015.
- [295] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, and G. O. Consortium, "Gene Ontology: tool for the unification of biology," *Nature Genetics*, vol. 25, pp. 25–29, May 2000.
- [296] C. Kumar and M. Mann, "Bioinformatics analysis of mass spectrometry-based proteomics data sets.," *FEBS Letters*, vol. 583, pp. 1703–1712, June 2009.
- [297] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.," *Nucleic Acids Research*, vol. 37, pp. 1–13, Jan. 2009.
- [298] A. Schmidt, I. Forne, and A. Imhof, "Bioinformatic analysis of proteomics data.," *BMC systems biology*, vol. 8 Suppl 2, no. Suppl 2, p. S3, 2014.
- [299] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.," *Nature Protocols*, vol. 4, no. 1, pp. 44–57, 2009.
- [300] R. P. Huntley, T. Sawford, M. J. Martin, and C. O'Donovan, "Understanding how and why the Gene Ontology and its annotations evolve: the GO within UniProt.," *GigaScience*, vol. 3, p. 4, Mar. 2014.
- [301] T. Cokelaer, D. Pultz, L. M. Harder, J. Serra-Musach, and J. Saez-Rodriguez, "BioServices: a common Python package to access biological Web Services programmatically," *Bioinformatics*, vol. 29, no. 24, pp. 3241–3242, 2013.
- [302] M. A. Harris, J. Clark, A. Ireland, J. Lomax, M. Ashburner, R. Foulger, K. Eilbeck, S. Lewis, B. Marshall, C. Mungall, J. Richter, G. M. Rubin, J. A. Blake, C. Bult, M. Dolan, H. Drabkin, J. T. Eppig, D. P. Hill, L. Ni, M. Ringwald, R. Balakrishnan, J. M. Cherry, K. R. Christie, M. C. Costanzo, S. S. Dwight, S. Engel, D. G. Fisk, J. E. Hirschman, E. L. Hong, R. S. Nash, A. Sethuraman, C. L. Theesfeld, D. Botstein, K. Dolinski, B. Feierbach, T. Berardini, S. Mundodi, S. Y. Rhee, R. Apweiler, D. Barrell, E. Camon, E. Dimmer, V. Lee, R. Chisholm, P. Gaudet, W. Kibbe, R. Kishore, E. M.

- Schwarz, P. Sternberg, M. Gwinn, L. Hannick, J. Wortman, M. Berriman, V. Wood, N. de la Cruz, P. Tonellato, P. Jaiswal, T. Seigfried, R. White, and Gene Ontology Consortium, "The Gene Ontology (GO) database and informatics resource.," *Nucleic Acids Research*, vol. 32, pp. D258–61, Jan. 2004.
- [303] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society. Series B. Methodological*, vol. 57, no. 1, pp. 289–300, 1995.
- [304] L. S. hle and S. Wold, "Multivariate analysis of variance (MANOVA)," *Chemo-metrics and Intelligent Laboratory Systems*, vol. 9, no. 2, pp. 127–141, 1990.
- [305] W. J. Conover and R. L. Iman, "Rank Transformations as a Bridge Between Parametric and Nonparametric Statistics," *American Statistician*, vol. 35, no. 3, pp. 124–129, 1981.