DISSERTATION ZUR ERLANGUNG DES DOKTORGRADES

DER FAKULTÄT FÜR CHEMIE UND PHARMAZIE

DER LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

DEVELOPMENT AND APPLICATION OF QUANTITATIVE PROTEOMICS STRATEGIES TO ANALYZE MOLECULAR MECHANISMS OF NEURODEGENERATION

Daniel Hornburg

aus

Magdeburg, Deutschland

2015

Erklärung

Diese Dissertation wurde im Sinne von §7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Matthias Mann betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe von mir erbracht.

München, 21.05.2015

_____

Daniel Hornburg

Dissertation eingereicht am 22.05.2015

1. Gutachter: Prof. Dr. Matthias Mann
2. Gutachter: Prof. Dr. Franz-Ulrich Hartl

Mündliche Prüfung am 06.08.2015

Summary

Neurodegenerative disorders such as Alzheimer's disease, Huntington's disease, Parkinson's disease, Amyotrophic Lateral Sclerosis or Prion diseases are chronic, incurable and often fatal. The risk to develop neurodegeneration increases with age, making these pathologies especially devastating in aging industrialized societies. Neuronal death leads to functional impairment of the affected region. A cardinal feature of all neurodegenerative disorders is the accumulation of misfolded and aggregated proteins. Depending on the disease, these aggregated proteins are cell type specific and have distinct cellular localizations, compositions and structures. Despite intensive research, the contribution of protein misfolding and aggregation to the cell type specific toxicity is not completely understood. The increasing availability of "omics" approaches opens up novel strategies to characterize the physiological properties of neurons and to investigate pathological changes in the context of neurodegeneration. In recent years, quantitative proteomics has matured into an exceptionally powerful technology providing accurate quantitative information on almost all cellular proteins as well as protein interactions, modifications, and subcellular localizations, which cannot be addressed by other omics technologies.

The aim of this thesis is to investigate key features of neurodegeneration such as misfolded proteins and toxic protein aggregates with cutting edge proteomics, presenting a technological "proof of concept" and novel insights into the (patho)physiology of neurodegeneration. In *article 1*, we analyzed the complete proteome of primary motoneurons, providing a reference protein inventory for these specialized cells. Such data can help to understand the physiological properties of these cells and their vulnerabilities causing specific pathologies. Furthermore, we compared primary neurons to neuronal model cell lines commonly used in research. We provided quantitative measures on proteomic differences, but also demonstrated that some disease relevant proteins and pathways are unchanged. Based on these data, the suitability of cellular model system for motoneuron studies can be evaluated.

Cells have evolved sophisticated mechanisms to maintain homeostasis of the proteome (proteostasis). Chaperones facilitate folding and can rescue partially unfolded proteins. Misfolded proteins and aggregates are usually degraded by the ubiquitin proteasome pathway and autophagy. Failure of proteostasis causes accumulation of protein aggregates and is a feature of almost all neurodegenerative disorders. However, a comprehensive understanding of how aggregates can exert toxicity is still missing. In *article 2*, we designed a proteomics approach to investigate the quantitative composition of Amyotrophic Lateral Sclerosis associated aggregates composed of poly-dipeptide

repeats. These aggregates displayed strong association with many different proteins including a significant part of the ubiquitin / proteasome system. Moreover, we identified new disease relevant proteins such as the cargo adapter protein Unc119. Characterization of Unc119 function revealed that its sequestration causes neuron specific toxicity.

In order to gain new insights into the cellular responses to toxic protein aggregates, we investigated huntingtin aggregates in yeast. In *article 3*, Hartl and coworkers first identified proteins in yeast that modulate huntingtin toxicity. We set out to quantify changes in the aggregate composition by proteomics and correlated these changes with the toxicity. We showed that a prion-like domain of the endogenous protein GTS1p reduces huntingtin aggregate toxicity by altering the aggregate structure and composition. It also increases accessibility for chaperones while preventing the aggregate from sequestering other cellular factors. These findings suggest novel cellular strategies to cope with proteostasis stress.

Alzheimer's disease is the most prominent neurodegenerative disorder. The accumulation of extracellular protein aggregates, amyloid Aβ plaques, and intercellular amyloid tangles are cardinal features of Alzheimer's disease. The proteolytic pathway in which Aβ is generated from the amyloid precursor protein by the γ- and β-secretase has been extensively studied. In theory, the inhibition of β-secretase should prevent production of neurotoxic peptides, however, studies on β-secretase knockout mice revealed the development of neurological defects [23]. In a collaborative effort with the group of Christian Haass we identified a novel proteolytic pathway of the amyloid precursor protein. Employing high performance mass spectrometry instrumentation and an optimized proteomic workflow, we identified and characterized a so far unknown neurotoxic peptide, termed Aη-α (*article 4* in press). This finding provides a new perspective on Alzheimer's disease pathology. Intriguingly, the production of neurotoxic Aη-α competes with β-secretase activity, which at least partially explains the neurological defects in β-secretase knockout mice and is highly relevant for drug development.

Cutting edge proteomics technology today allows the identification of more than 10,000 proteins in individual cell types (*article 1*). However, proteomics still greatly benefits from increased mass spectrometer performance in terms of number of identified peptides and quantification accuracy. We evaluated the performance of the latest generation of hybrid quadrupole Time-of-Flight and quadrupole Orbitrap mass spectrometers in *article 5* and *article 6*. For the Orbitrap we assessed the complexity of the parameter space that affects machine performance by applying mathematical models. These data allowed us to determine the optimal settings and to maximize mass spectrometer performance.

Table of Contents

"Any time scientists disagree, it's because we have insufficient data. Then we can agree on what kind of data to get; we get the data; and the data solves the problem. Either I'm right, or you're right, or we're both wrong. And we move on. That kind of conflict resolution does not exist in politics or religion."

Neil deGrasse Tyson [24]

Introduction

1. Neurodegeneration

"The human brain has 100 billion neurons, each neuron connected to ten thousand other neurons. Sitting on your shoulders is the most complicated object in the known universe."

Michio Kaku [25]

1.1. The Neuronal Network

The human brain is one of the most complex systems known. More than 86 billion neurons are estimated to structure a network controlling most body functions and reflecting our consciousness [26]. The connectivity of neuronal circuits exhibits enormous plasticity during development and to a lower extent also in adults. Neurons themselves are very specialized and morphologically distinct cells. Motoneurons, for instance, are extremely elongated cells that connect the brain with muscles of the body. Neurons are highly connected to each other via cellular extensions called dendrites, which receive stimuli from synapses of other neurons. Starting with only a few at birth, some neurons eventually develop up to 100,000 dendrites, providing thousands of dynamic synaptic connections to other neurons, eventually reaching a network complexity of 100 trillion connections [27-29]. Within this neuronal network, information is electrically and chemically coded and transmitted. This dynamic connectivity in the brain is one of the key elements in cognitive processes such as learning and memory formation.

A 90 years old person will not outrun a 20 years old person but might well beat her in chess. Rodriguez and coworkers pointed out that the brain is one of the few organs that can remain functional even in old age [30]. Specialized non-neuronal cells, termed glial cells, maintain the homeostasis of the neuronal network over decades. Interestingly, glial cells have at least similar abundance as neurons [26]. Some glial cells such as the oligodendrocytes (or Schwann cells in the peripheral nervous system) insulate neurons with myelin sheaths, providing metabolic and structural support while increasing efficiency of neuronal signaling along axons [30]. Astroglia play a central role

in modulating the synaptic activity as well as neuronal crosstalk [11]. Glial cells also protect against infection and tissue malfunction. In particular microglia, which essentially are brain-resident macrophages, provide a pro- or anti-inflammatory environment [18]. Their limited capacity for regeneration and their unique metabolic, functional and morphological features render neurons particular vulnerable. Deterioration of homeostasis in the nervous system is associated with a spectrum of neurodegenerative disorders.

Glia Cells and Neuronal Signaling

The initial assumption that glia cells do not directly participate in neuronal signaling was challenged in studies performed on astrocytes, the most common type of glia cells in the brain. These demonstrated the capacity of astroglia to release gliotransmitters (ATP or glutamate), which modulate synaptic activity and trigger excitatory or inhibitory responses in the neurons [11].

1.2. Pathological Events in Neurodegeneration

Local and global neuron death leads to functional impairment of the affected region and causes incurable and often fatal disorders such as Alzheimer's disease (AD), Huntington's disease (HD), Parkinson's disease (PD), Amyotrophic Lateral Sclerosis (ALS) or Prion diseases. These diseases are associated with a spectrum of pathological changes on the cellular level, for many of which the question of cause or consequence is not yet solved.

The hallmark of all neurodegenerative disorders (ND) is accumulation of misfolded and aggregated proteins (Table 1). This accumulation and the risk of developing a ND both increase with age, rendering it an increasing burden for industrialized societies [31-33]. The protein aggregates often have disease specific properties such as composition, localization and structure. Some of these aggregates, however, occur in multiple pathologies. For instance, alpha synuclein inclusion are characteristic for PD, but can also be detected in 50 % of all AD cases [34]. It is well established that aggregates and misfolded oligomers can exert toxicity in cells [35,36]. Apart from this direct toxicity, consequent decline of proteostasis can further promote accumulation of misfolded proteins. Being both cause and consequence renders protein misfolding a catastrophic self-promoting cascade [37].

Major Causes of Death

The life expectancy in developed countries has doubled in the last two centuries, which renders age related pathologies a growing problem [1,2]. NDs are now among the major age related causes of death. In contrast to many other diseases, neurodegeneration associated fatalities are still rising and the WHO projects that in 2050, 106 million people will suffer from NDs [6,7].

The risk of developing ND is associated with a variety of genetic mutations. In the case of HD, mutations in a single gene resulting in an elongated polyQ stretch of the huntingtin protein (HTT) are causing the pathology. For other NDs, alterations in more than one risk allele have been reported. Mutations in the amyloid precursor protein APP or its processing enzyme PSEN1 can lead to an early onset AD and dozens of mutations are associated with ALS and FTD [38,39]. Neurodegeneration is accompanied by a deterioration of several cellular processes, including the proteasome / ubiquitin system (UPS), oxidative stress response, neurotransmission (excitotoxicity), mitochondrial function, and RNA metabolism and axonal or dendritic transport (Figure 1) [40]. In addition, a growing body of evidence points towards a disturbed cellular communication in neurodegeneration. Astroglia degeneration and altered neuronal crosstalk, as well as inflammatory processes can weaken neuronal homeostasis [30,41]. Microglia can be activated by protein aggregates and an impairment of the blood
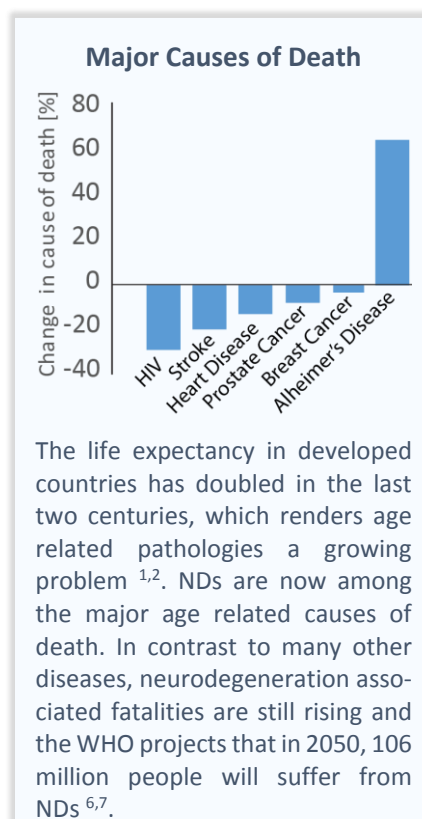
brain barrier can amplify the neuronal decay by promoting neuroinflammation [42,43]. Especially these inflammatory processes are considered a major driver of widespread neuronal death.

Taken together, multiple factors contribute to the development and progression of neurodegenerative disorders. In 2007 Sulzer suggested a "*Multiple hit hypotheses*" for PD in analogy to cancer [44]. This assumption probably also holds true for other NDs where genetic and environmental factors modulate disease onset and progression.
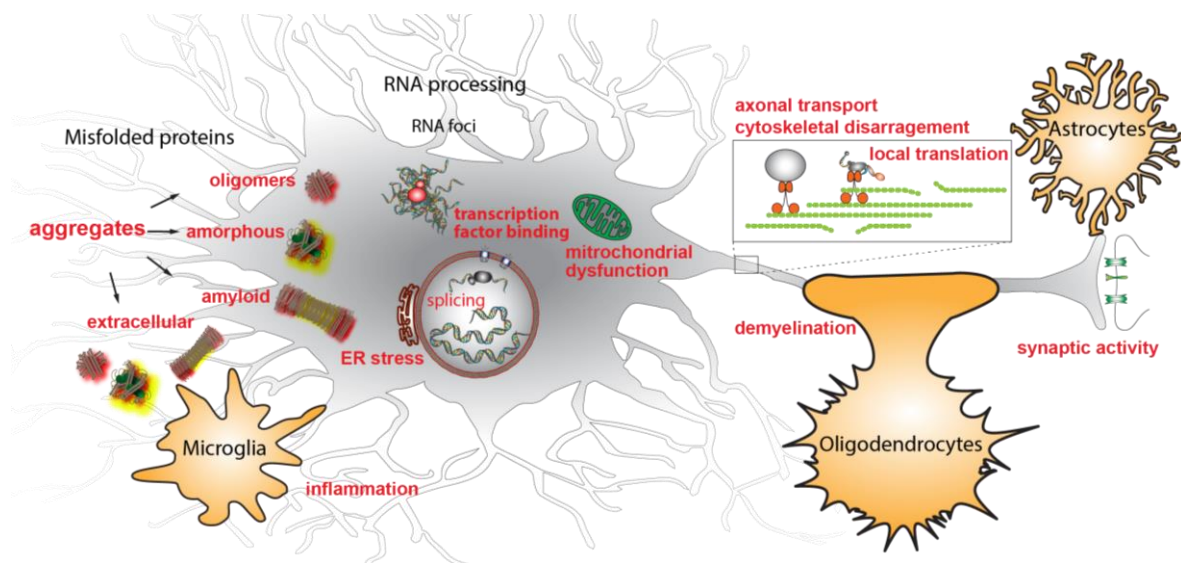
Figure 1 Pathological alterations in neurodegeneration: Apart from the accumulation of protein aggregates, neurodegenerative disorders involve a variety of intra- and intercellular processes (depicted in red). The discrimination of primary and secondary effects is usually difficult due to slow disease progression before manifestation of symptoms [18,30,34,40,41,45].

Inflammation in Neurodegeneration

Inflammation is essential to maintain homeostasis by removing pathological stimuli and by promoting the healing process. If dysregulated, inflammation is deleterious especially for tissues with limited regenerative capacity such as the central nervous system (CNS). Under physiological conditions, the CNS is an immune privileged organ and therefore insulated from many activities of the immune system. It lacks connection to the lymphatic system and the blood-brain barrier shields the CNS from immunoglobulin and immune cell infiltration. Resident microglia remove cell debris, clear off neurotropic pathogens and tumors but in contrast to their blood-circulating counterparts provide a rather anti-inflammatory environment [18].

Table 1 Major NDs with characteristic protein aggregates [46,47]

Diseases	Pathology	Aggregates	Location
Alzheimer's disease[46]	whole brain, mental impairment	TAU	Cytoplasm
		Amyloid Beta (Aβ)	Extracellular
Creutzfeldt–Jakob disease[46]	whole brain, mental impairment	Prion protein (PrP^{sc})	Cytoplasm
			Extracellular
Polyglutamine expansions (including Huntington's disease or spinocerebellar ataxia)[46]	whole brain, mental impairment, behavioral changes	PolyQ proteins	nuclei and cytoplasm
ALS and Frototemporal dementia [47,48]	motoneurons and temporal and prefrontal cortex, paralysis and mental impairment	SOD1, TDP43, poly dipeptide repeat	nuclei and cytoplasm
Parkinson's disease [34]	substantia nigra, motorsymptoms	Lewis bodies (α-synuclein)	cytoplasm

1.3.   Amyotrophic Lateral Sclerosis

ALS is a rare and fatal ND of the motorsystem, affecting 1 in 50,000 people. Degeneration of upper and lower motoneurons in the brain stem and spinal cord as well as in the motor cortex lead to progressive and ultimately complete paralysis. Most patients die within three to five years, although in rare cases can survive much longer [47]. Death is usually caused by loss of neurons that innervate breathing. Recent evidence suggests that ALS is at one extreme in a disease continuum whose other extreme is frontotemporal dementia (FTD), a pathology characterized by the progressive loss of neurons in the temporal and prefrontal cortex causing mental retardation and behavioral changes [49].

As with most NDs, one hallmark of ALS is the accumulation of misfolded proteins in the cytosol and nucleus. These aggregates have different properties depending on the genetics of the pathology. The number of genes that contribute to ALS pathology is very large compared to other NDs and currently the ALSOD database lists 126 ALS-related genes (2015/21) [38]. Strikingly, genetic causes are only identified for 10 % of all ALS cases. This could be an underestimation due to incomplete knowledge of the disease genetics [50]. The corresponding proteins are involved in diverse cellular processes such as reactive oxygen species detoxification (SOD1), RNA metabolism (TDP-43, FUS), the cytoskeleton (DCTN1, MAPT) or proteostasis (UBQLN2). Some of the ALS cases are associated with loss of function mutations, but most are linked to gain of function mutation [47]. Although the specific pathological manifestations of ALS suggest a common link between the associated genes, a common ALS disease mechanism remains elusive. Progression of ALS is promoted by neuroinflammation, which is mediated by microglia switching from immune suppressive to pro-inflammatory phenotype and T-cell invasion [18].

In 2011 two groups independently reported the most common genetic cause for ALS and FTD [51,52]. The associated hexanucleotide expansion in an intronic sequence accounts for 40 % of the inherited instances of ALS and was also found in 7 % of all sporadic (no family history) ALS cases [53]. The physiological function of the c9orf72 protein is unknown and three different pathological mechanisms have been proposed. First, the repeat expansion alters the splicing of the c9orf72 protein, which might lead to a haploinsufficiency [54]. A similar mode of action has been described for mutations in progranulin (PGNR), which are causing FTD [55,56]. Second, the hexanucleotide repeat expansion is transcribed, leading to high amounts of mRNA that accumulate in RNA foci. These RNA foci have been suggested to be toxic by sequestering RNA binding proteins, similarly as described for myotonic distrophy type 2, where the pathology is linked to RNA repeats that act as a sink for splicing factors [54,57]. Third, hexanucleotide repeat associated peptide aggregates might themselves cause toxicity. In support of this possibility, two groups reported a start-codon

independent translation, termed 'RAN' translation [58], of the hexanucleotide repeat resulting in different species of aggregation prone dipeptide repeats [48,59].
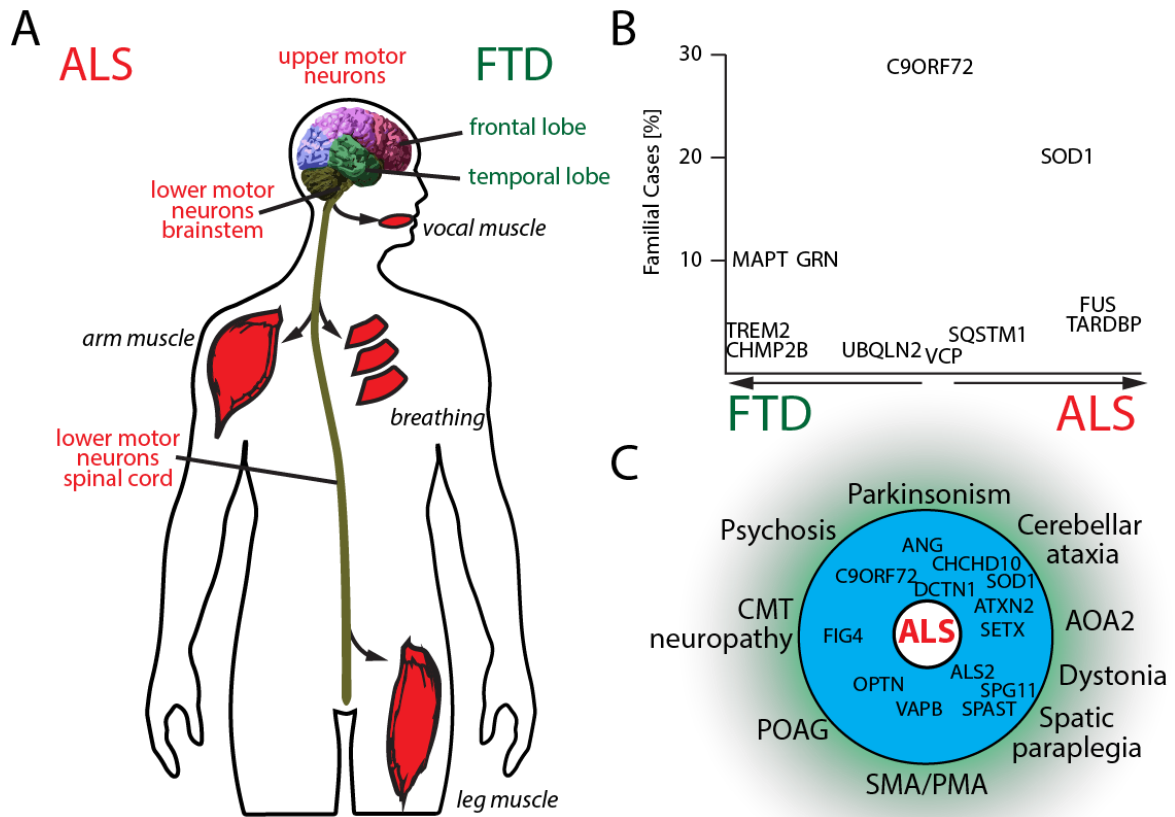
Figure 2 ALS and FTD pathology: A) FTD and ALS comprise the extremes of a common disease spectrum. ALS primarily affects the motorcortex, motoneurons in the brain stem as well as lower motoneurons in the spinal cord. In contrast, FTD is characterized by increasing degeneration in the frontal and temporal lobe [47] (adapted [60]). B) Various genes cause ALS and FTD pathology. While mutations in some genes specifically lead to either ALS or FTD pathology, the most common genetic cause c9orf72 is shared by both (adapted [61]). C) For ALS, neurological pathology depends on the genetics. Mutation in some genes tend to cause motor phenotype while some are also associated with other symptoms (adapted [61]) (CMT: charcot-marie-tooth disease; POA: primary open-angle glaucoma; SMA / PMA: spinal muscular atrophy / progressive muscular atrophy; AOA2:ataxia with oculomotor apraxia type 2)

1.4. Expansions of Repetitive Genetic Elements in Neurodegeneration

Expansions of repetitive genetic elements, termed microsatellite expansions, are associated with various NDs such as above described c9orf72 associated ALS and FTD [62], HD, spinal bulbar muscular atrophy, and spinocerebellar ataxia types 1, 2, 3, 6, 7, and 17 or fragile X-associated tremor ataxia syndrome. These repeats were traditionally termed either coding or noncoding repeats, depending whether or not they are located within an ATG-initiated open reading frame. Among those, poly glutamine (polyQ) associated pathologies comprise a prominent group of repeat associated neurological disorders with toxic gain-of-function on the protein level (see chapter 1.6.2). Noncoding repeats can also mediate toxicity on the transcriptional level where the RNA acts as a sink for RNA binding proteins. Beside these two mechanisms, repetitive genetic elements can prime toxicity via repeat associated non-ATG translation ('RAN' translation [62]).

In most cases, mRNA translation in eukaryotes is initiated by the AUG start codon. However, this canonical translation does not cover the whole spectrum of proteins and many instances of alternative translation initiations have been reported [63]. In 2011 a novel mode for AUG independent translation of microsatellite repeat expansions was reported, which does not initiate via Met-tRNA [64]. A RAN translation to polyQ for repetitive CAG but not CAA sequences was observed, which indicates that initiation depends on the capacity to form RNA hairpins [64]. This hairpin mediated translation mimics the initiation tRNA with a RNA stem loop structure and is mechanistically similar to that at internal ribosome entry sites used by certain viruses [58]. RAN translation occurs in several microsatellite disorders such as spinocerebellar ataxia type 8 or FTD and c9of72 pathology of ALS [48,59,64]. In RAN translation, all three reading frames are translated. The resulting proteins can vary in abundance, which might be due to different translation efficiency or protein stability. The efficiency of RAN also depends on the repeat length. A minimum length can be required and simultaneous translation on very long repeats can increase the amount of the corresponding protein [64].

1.5. Protein Folding

1.5.1. Protein Structure and Function

Despite the diversity of NDs, protein misfolding and aggregation are cardinal features. The functionality of proteins depends on the three-dimensional shape that emerges from the folding of the poly-amino acid sequence. The folding of proteins is a complex and yet not fully understood process. In the 1960s Anfinsen and co-workers performed refolding experiments and hypothesized that all the information for the correct tertiary structure is encoded in the primary structure of the polypeptide allowing it to *de novo* fold into the native conformation [65]. Even for small polypeptides its degrees of freedom result in a combinatorial explosion that would make a stochastic folding process impossible. Instead of pure stochastic path, the folding therefore has to happen in a more guided manner. Several groups suggested a funnel-like energy landscape shaped by the thermodynamic properties of all possible conformations [66]. This energy landscape would facilitate a guided folding process within the Anfinsen assumption. In addition, the environment of the protein affects the thermodynamics and kinetics and the stable conformations and folding pathways can be remodeled and controlled. Enzymes facilitating cis-trans isomerization of proline (peptidyl-prolyl isomerases) and disulfide isomerases can effectively reduce kinetic barriers in the folding process [67,68]. For small proteins (100 amino acids), folding usually happens on a sub-second time scale and involves interaction of key residues and reduction of hydrophobic surfaces [69]. In contrast, for larger proteins this process is more complicated with folding taking place simultaneously at different parts of the protein and involving multiple folding intermediates [70,71]. Those intermediates can promote further folding but can also act as a trap for incorrect conformations [72,73].

The Levinthal Paradox

The Levinthal paradox is a thought experiment that shows that protein folding cannot be a random process [14]. The combinatorial explosion would not allow folding of even small proteins within observed time scales of seconds to minutes. Assuming three degrees of freedom per amino acid and a sampling rate of 10^{13} conformations per second, a 16mer samples all conformations in less than 2 µs, a 38mer takes one day, a 45mer a whole PhD period (three years) whereas a 66mer might not have been folded since the 'big bang'.
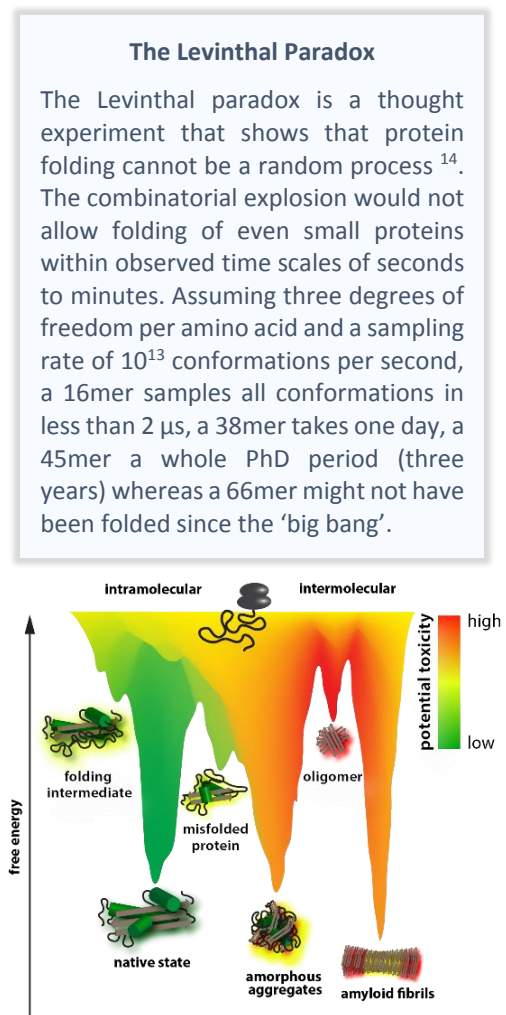

Figure 3 Protein conformation energy landscape: A simplified energy funnel that depicts the thermodynamic stability of protein conformations and kinetic barriers. Folding intermediated can either fold spontaneously or are assisted by chaperones. Protein misfolding or partial folding can promote potentially harmful amorphous or amyloid aggregates or oligomers (adapted [17]).

1.5.2.   Protein Quality Control

The assumption that all polypeptides directly fold into their native conformation only holds true for some small proteins or *in vitro* systems. A tendency to aggregate under certain conditions can be observed for a wide spectrum of sequences and it has been suggested to be an inherent property of polypeptide chains [74]. In order to function properly, proteins retain conformational flexibility after folding, rendering the native structure only marginally thermodynamically stable [17]. To facilitate and maintain correct conformation and prevent trapping of folding intermediates, cells have evolved a sophisticated set of quality control mechanisms. This proteostasis network consists of hundreds of proteins. *De novo* folding, refolding, oligomeric assembly, protein trafficking, disaggregation and proteolytic digestion is assisted by over 200 chaperones and cochaperones. In addition, roughly 600 proteins of ubiquitin / proteasome and autophagy system are responsible for removing misfolded, aggregated or obsolete proteins [17].

Intrinsically Disordered Proteins

Some proteins do not have a defined secondary and tertiary structure but contain major disordered regions. Those flexible interfaces are thought to participate in cellular signaling. The structural flexibility allows conformational adaption to different interaction partners but can make the protein specifically vulnerable to misfolding [16].

Chaperones assist proteins in a concerted manner to maintain or fold into the native state [75]. The general principle is to reshape the energy funnel transferring aggregation prone intermediates to the native state and to shield them from deleterious interactions [76]. Chaperones do not directly contribute conformational information and usually do not accelerate folding directly but instead promote correct protein folding indirectly by inhibiting competing processes. Chaperone families like HSP70 and the chaperonins HSP60 are multimeric molecular machines. They recognize hydrophobic sidechains and engage in co-factor and ATP dependent binding and release cycles. The first hurdle a polypeptide that belongs to a large domain (for instance 100 to 300 amino acids) faces, is that the nascent chain exiting the ribosome will not have all the native intramolecular moieties available for correct folding, but instead could favor non-native interactions *via* exposed hydrophobic regions. In eukaryotes, the ribosome-associated complex (RAC) and nascent polypeptide associated complex (NAC) are responsible for stabilizing the nascent chain. In prokaryotes, this process is facilitated by the trigger factor (TF). Downstream of this initial interaction the polypeptide can interact with HSP70s (DnaK in prokaryotes) and the co-chaperone HSP40 (DnaJ in prokaryotes), which inhibit premature misfolding. Some proteins require further assistance to fold correctly into their native state. This posttranslational process is mediated by chaperonins (HSP60 of mitochondria, GroEL in bacteria or TRiC in eukaryotic cytosol). Chaperonins are large double ring complexes that act downstream of HSP70s and can enclose translated proteins of up to 60 KDa in a cylindrical cage.

Chaperonins remove entropic and kinetic barriers in the folding path while increasing the energetic penalty for misfolding. The confinement provides a defined chemical environment and concentration and restricts steric freedom of the unfolded conformation. Complete folding of proteins can require multiple cycles of confinement and release, which is triggered by ATP hydrolysis. In contrast to HSP60s chaperonins, the TRiC does not require HSP10 (GroES in prokaryotes) and has substantially longer cycle times [17].

After successful folding, the cell has to assure surveillance of proteins. Random structural fluctuations in folded proteins or external perturbations, such as heat stress, can expose aggregation prone intermediates that can accumulate over time. In such cases, the chaperones shield partially unfolded proteins, which provides time for refolding while excluding detrimental effects on other proteins [77]. Highly abundant chaperones of the HSP90 family act downstream of the HSP70 system and assist in final maturation and stabilization of proteins. Among the HSP90 clients are numerous proteins involved in signal transduction making HSP90 a proteostasis hub for signaling pathways [17]. The unfoldase activity of HSP100 chaperones (bacterial ClpB, yeast Hsp104) contributes to refolding and degradation [78].

If the cellular rescue system for misfolded proteins fails, cells need to eliminate these proteins (Figure 4). For this purpose different protein clearance pathways have evolved [45]. One of these is autophagy, by which large aggregates are encapsulated and degraded in lysosomes. A related pathway, chaperone assisted autophagy, degrades individual proteins by a chaperone guided translocation to the lumen of lysosomes [79]. Moreover, the UPS can mediate protein degradation. The proteasome is a barrel-shaped protein complex that requires at least a partial unfolding of substrate proteins for their degradation. Finally, cells dilute aggregates via cell division [80].

1.6. Deterioration of Proteostasis

When the cellular folding and degradation machinery fails to maintain proteostasis, misfolded proteins and protein aggregates accumulate. Many factors promote the decline of the proteostasis network. Age decreases the capacity of proteasomal degradation pathways and of autophagy [81,82]. This inevitably increases the burden of aggregated proteins as shown in the *C.elegans* and *D.melanogaster* model systems [83-85]. A study performed in 2012 further underlined the detrimental effects of age related accumulation of misfolded proteins, by showing that knockdowns of aggregation prone proteins in *C.elegans* can result in increased lifespan [86]. Several studies link impaired autophagy to the reduced capacity to clear aggregates, eventually leading to neurodegeneration [87-90]. Moreover, aggregates themselves have been suggested to compromise proteostasis resulting in a "viscous circle" of aggregates and impaired protein folding [37].

Besides impaired cellular capacity to degrade and clear misfolded proteins, other factors contribute to accumulation of aggregated proteins. Remarkably, in some instances point mutations are sufficient to induce protein aggregation. This was demonstrated, for instance, for systemic amyloidosis, transthyretin amyloid pathologies and a luciferase aggregation sensor [91-93]. Moreover, the local environment affects protein conformation. Dialysis associated amyloidosis is a pathology in which very high protein concentrations and an altered environment by themselves change the aggregation propensity of the plasma protein beta-2-microglobulin (B2m) [94]. This is not surprising considering that the aggregation process is a higher order reaction and therefore concentration dependent, as the initial crystallization enthalpy needs to outweigh the entropic penalty [95]. Finally, chemical modifications, heat and cold stress can induce protein misfolding [96,97].
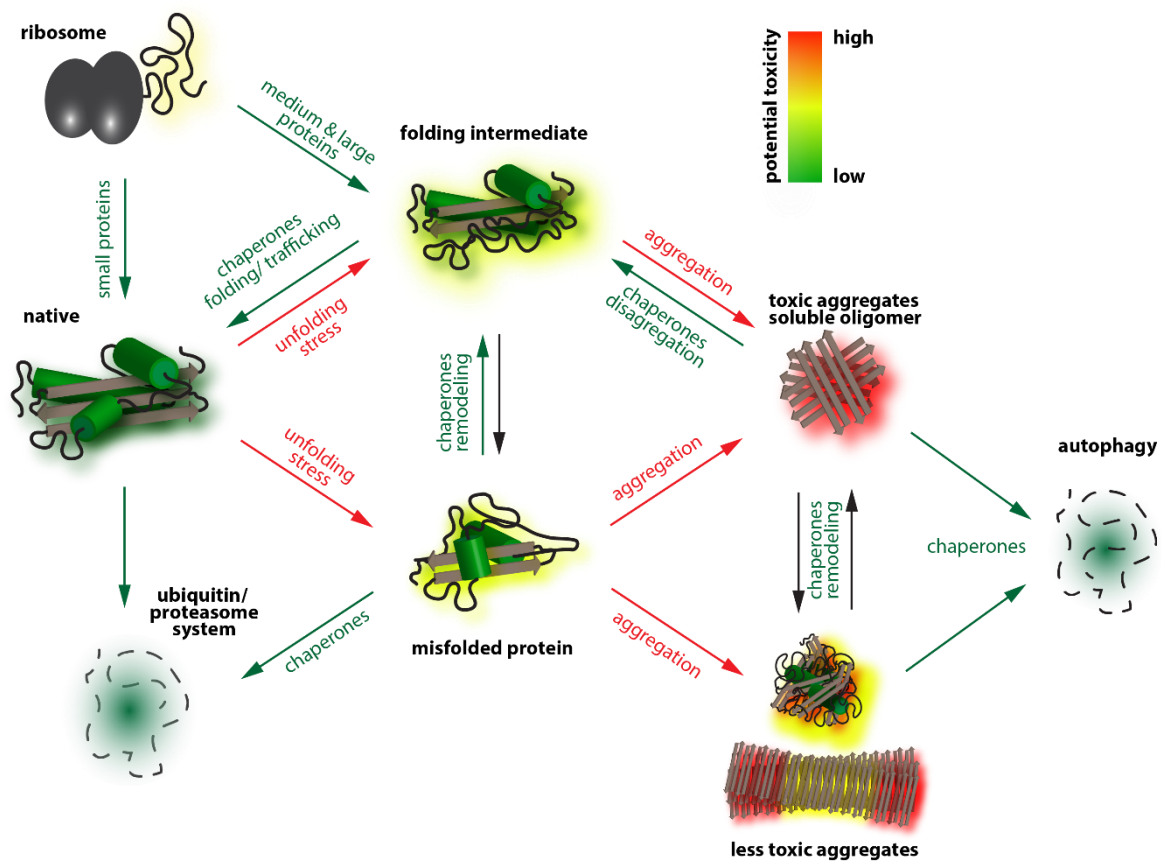
Figure 3 Protein life-cycle: Proteostasis is maintained by various cellular machineries. Misfolded proteins are rescued by chaperones or degraded via the ubiquitin proteasome system. Very large aggregates cannot be degraded in the proteasome and are decomposed by autophagy instead. Failure of the proteostasis system can lead to the accumulation of harmful misfolded proteins and is a general feature of neurodegeneration (adapted [37]).

1.6.1.   Misfolded Proteins

Proteostasis deficiency and formation of protein aggregates are associated with many diseases like type 2 diabetes, peripheral amyloidosis, lysosomal storage disease, cancer, cystic fibrosis and cardiovascular disease. However their most prominent role is in neurodegeneration [17]. Aggregates differ with regard to their localization within or outside the cell and their structural properties. Many aggregates form ordered amyloid fibrils of 6 to 10 nm with cross-β structure, in nucleation dependent polymerization [98-101]. In addition to highly ordered amyloid-sheets, amorphous aggregates can be observed. Beside these insoluble inclusion, misfolded proteins can also form soluble oligomers. Whether proteins adapt amyloid, amorphous or oligomeric structures depends on the amino acid sequence and the chemical and physical context [96].

Protein misfolding has a negative effect on cell viability [37]. Whether this is caused by toxicity of large aggregates or oligomeric intermediates is controversial and may depend on the circumstances [101,102]. Compared to oligomers large aggregates have a relative small surface to mass ratio and thereby a small interface to the cell. Therefore, creating very large aggregates can be an effective way to reduce the exposure of misfolded proteins. However, the reduced interface also compromises accessibility for the cellular clearance machinery. In sum, the toxicity of soluble oligomers seems to be higher than that of large aggregates [37,103]. Many mechanisms have been proposed by which protein aggregates and oligomers mediate toxicity. In contrast to 'loss of function' by removal of the aggregating protein itself, 'gain of function toxicity' can induce toxicity *via* multiple routes. For instance, aggregated proteins might cause inflammation, deteriorate membrane integrity and compartmentalization, gradually compromise the proteostasis machinery or sequester essential cellular factors into the aggregate [35,104,105].

1.6.2. Poly Glutamine Toxicity

Despite the general capacity to misfold, some sequence features such as polyQ stretches render proteins specifically aggregation prone. Several inherited polyQ associated pathologies are known: SCA type 1, 2, 3, 6, 7 and 17, dentatorubal-dallidoluysian, spinobulbar muscle atrophy and HD [106]. Among these nine, HD is the most prominent one. HD is a neurodegenerative disorder that impairs motorcontrol, causes behavioral changes and mental decline. In HD patients, the expansion of a CAG repeat in exon 1 of the huntingtin gene translates to an aggregation prone and toxic protein containing an elongated polyQ stretch [107]. Normal alleles of HTT usually contain 16 to 20 repeats, while more than 35 repeats are associated with high penetrance of HD [106]. The elongated polyQ stretch in HD leads to the deposition of aggregated htt in the cytosol and nucleus. These aggregates are β-sheet rich, highly ordered amyloid fibers that interact with hundreds of proteins [108,109]. The sequestration of many different proteins could explain why HD, despite being a monogenetic disease, causes a wide set of cellular dysfunctions, like transcriptional dysregulation, impaired protein degradation, mitochondrial dysfunction, disrupted neuronal signaling and impaired protein folding [106]. HTT is not only relevant from a pathological point of view but offers a good model to study the effects of toxic protein aggregates. Chaperones such as HSP70 and TRiC can specifically block formation of fibrillary amyloid HTT aggregates in yeast. HSP70 seems to interfere with initial nucleation, allowing TRiC binding formation of less toxic amorphous HTT aggregates [110,111].

1.6.3. 'Contagious' Protein Misfolding in Prion Diseases

The central role of protein misfolding in neurodegeneration is underlined by transmissible prion disorders. Prion-disorders share many characteristics with other NDs including brain vacuolation, astrogliosis, spreading neuronal death and protein aggregation [112]. However, they differ by being transmissible and are referred to as transmissible spongiform encephalopathies (TSEs). This transmission was observed in nature and can be performed in vitro across some species. Vectors for the misfolding information is the prion protein PrP^{SC} that amplify this information in an autocatalytic fashion (Figure 5). The paradigm example of a TSE is bovine spongiform encephalopathy (BSE) and its associated human pathology, the variant Creutzfeldt-Jakob disease (vCJD) [113,114]. Other prominent prion diseases are Scrapie in sheeps and Kuru in humans. Kuru was mainly transmitted via cannibalism in the highlands of Papua New Guinea [115].

Prions are epigenetic elements in that they can carry and store information without directly using nucleic acids. Prions can exist in different strains, outcompete each other during propagation and thereby undergo natural selection [116-118]. Instead of always being toxic, such a system of replicating information could potentially be employed functionally by the cell. In yeast, many proteins have been annotated as potential prions [119], for instance, the [PSI+], which is a prion form of the translational release factor SUP35. Although [PSI+] is almost exclusive to yeast lab strains, positive effects mediated by this prion have been suggested. The prion conversion to [PSI+] causes a certain degree of read-through capacity during translation that compensates nonsense codon mutations and can result in C-terminal elongated proteins [120]. *Article 3* shows a paradigm example of how prions in yeast help the cell to survive proteostasis stress by associating with otherwise toxic aggregates. Intriguingly, many proteins on higher organism also have prion like domains [121].
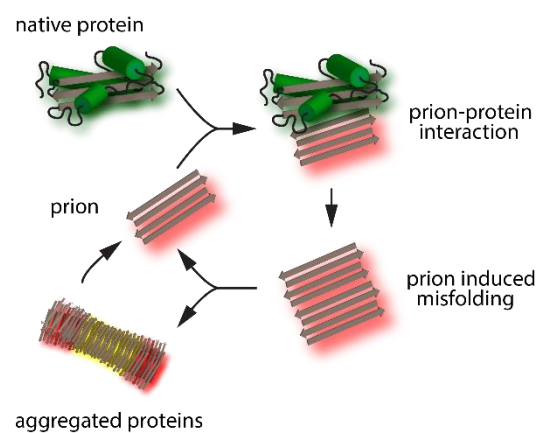

Figure 4 (Auto)catalytic propagation of prions: Prions can propagate and amplify folding information. Their high physical and chemical resistance is probably one of the reason for the contagiousness of prion-associated pathologies like Kuru or BSE.

Unraveling the Identity of Prions

In the 1960s an unusual high resistance to irradiation and ionization and very small size of the pathological agent in Scrapie suggested a so far unknown pathological entity [3,4]. In 1967, Griffith suggested an autocatalytic replication mechanism for a toxic protein [9]. Such a protein was termed prion (proteinaceous infectious particle (Greek *on* = 'being')) by Stanley Prusiner five years later [19]. This protein hypothesis was substantiated by showing that affinity enrichment of the prion protein (PrP) allows isolation of the transmissible agent [20]. In 1993 Brueler et al. demonstrated that mice lacking the PrP gene are resistant to Scrapie infection in contrast to the wild type. However, it took another 14 years to finally prove the "protein only" hypothesis by *de novo* generation of infectious prions converting native PrP^C to the toxic PrP^{SC} [22].

1.7.   Central Role of the Proteome

Impaired proteostasis is the cardinal feature of NDs. Knowledge about the proteome and its pathological disturbance before and during neurodegeneration might therefore be one of the most promising steps in understanding NDs and in providing a rational basis for prevention, early diagnose and treatment. However, analysis of the proteome is exceptionally challenging because it is made up of thousands of proteins with vastly different abundances and diverse chemical properties. Furthermore, proteins cannot be amplified like oligonucleotides. However, mass spectrometry (MS) is potentially a universal and extremely specific detection technology for proteins. Recent improvements in MS and associated proteomics technologies now provide a novel tool set for a comprehensive and unbiased view on diseases such as neurodegeneration as will be explained in the chapters below.

2. Quantitative Mass Spectrometry Based Proteomics

"...the constantly evolving gravimetric balance has been a faithful servant of the laboratory chemist and has played a major role in developing the analytical methods that are the foundation of contemporary chemical science. Perhaps the ultimate stage in the evolution of that balance is represented by the modern mass spectrometer. It is able to determine with high precision the masses of individual atoms and molecules by transforming them into ions and measuring the response of their trajectories in vacuo to various combinations of electric and magnetic fields."

John B. Fenn 2002 Nobel Laureate in Chemistry [122]

The central dogma of molecular biology states that the information is passed from DNA to RNA and is finally translated into proteins. Although there are exceptions to this unidirectional path, in most cases proteins are the functional entities in cells, working as enzymes, carriers or structural elements. The analysis of proteins has always been challenging due to their complex chemical nature and the absence of amplification procedures. Determining the mass of proteins using a mass spectrometer (MS) is one way to characterize them. Efficient ionization is a prerequisite for MS analysis. For decades, the fragile nature of proteins was not compatible with existing ionization techniques. In the late 80s, two gentle ionization techniques, electrospray ionization (ESI) and Matrix-assisted laser desorption / ionization (MALDI), paved the way to finally analyze proteins [123,124]. These technologies earned the Nobel Prize and proved to be the nuclei for the emerging field of proteomics. One of the advantages of ESI compared to MALDI is that proteins or peptides are ionized from the liquid phase. Therefore, ESI can be directly coupled to liquid chromatography and allows the analysis of complex protein and peptide mixture [125,126].

By far the most common way used to analyze complex protein mixtures is the bottom up approach (shotgun proteomics). Shotgun proteomics uses enzymes to digest proteins into specific peptides, which are then individually analyzed. Subsequently, downstream bioinformatics analysis reassembles the proteins from these peptides. In contrast, top-down proteomics subjects complete proteins to MS analysis. This has the advantage in principle that the entire protein with its specific combination of processing and modification events is analyzed together. This can be beneficial for

investigating isoforms or posttranslational modifications (PTMs) but is so far mainly limited to purified proteins. This is because the chromatographic separation is much more challenging and high molecular weight and complex charge patterns of the molecules make the MS spectra more difficult to obtain and to interpret [127].

2.1.   The Proteome

With slightly more than 20,000 genes, *H. sapiens* is placed half way between *G.gallus* (chicken) and *V.vinifera* (grape) in terms of genome size [128]. It is apparent that the number of genes is not indicative for the "complexity" of these organisms. Moreover, each cell within an organism contains the same set of genes but can show a completely different functionality. Physiological differences therefore have to originate from regulation of these genes, their translation into proteins, modification, localization and complex formation of these proteins. How many proteins individual human cells express at any time point is cell type specific and can exceed 10,000 [129]. Taking into account PTMs and splice variants increases the complexity of the proteome dramatically [130,131]. With MS based proteomics, different layers of proteomes can be characterized. Conceptually most straightforward is the investigation of complete proteomes of cells or tissues. Such studies aim for high depth (>10,000 proteins) to characterize the complete system [132,133]. Biological functions that involve protein-protein interactions, substoichometric protein modifications or specific cellular localizations, cannot be addressed with a complete proteome approach. In such cases, isolation and enrichment strategies are employed during sample preparation. Examples for subproteomes are protein aggregates characterized in *article 2* and *3*, the class of plasma membrane proteins such as receptors, secreted proteins, vesicular proteins or class of proteins bearing specific PTMs [134-137].

2.2.   Interaction Proteomics

Many proteins mediate their functions by interaction with other proteins, forming stable complexes or engaging in transient interactions. To study these interactions, classically proteins were isolated via affinity purification and interacting proteins were detected using Western Blots techniques. This targeted approach requires prior hypotheses to test for suspected interacting proteins and is only semi-quantitative. Common technologies like yeast two hybrid and affinity purification coupled to MS (AP-MS) have paved the way to high-throughput screening [138,139]. Initially, low resolution MS only allowed to identify but not quantify proteins. Therefore, sample preparation had to be sufficiently stringent to avoid background binding of proteins, which could compromise proper

identification of prey proteins. The increasing availability of quantitative MS facilitated a new strategy that reveals protein-protein interactions by quantitative enrichment [140]. With a quantitative readout, proteins binding nonspecifically can be discriminated from preys and used as internal control in form of stable background. This allows affinity enrichment instead of (harsh) purification and offers benefits such as preservation of near physiological conditions and the detection of weak interactors [141]. In *article 2* and *3* we employed such an approach to quantify proteins interacting with insoluble ALS associated aggregates in primary neurons and the huntingtin interactome in yeast.

2.3.   From Cells to Proteomes

2.3.1.   Sample Preparation

Shotgun proteomics requires the proteolytic digestion of proteins to peptides and the removal of any agents that are incompatible with MS. Over the years, many different lysis, enrichment and fractionation protocols have been developed. Currently, there is not one approach, which fits all purposes and the decision which method to select depends on criteria including the amount of material, the source (tissue, cell culture), the biological question or the quantification method to be employed. In the following, the major steps in sample preparation are described.
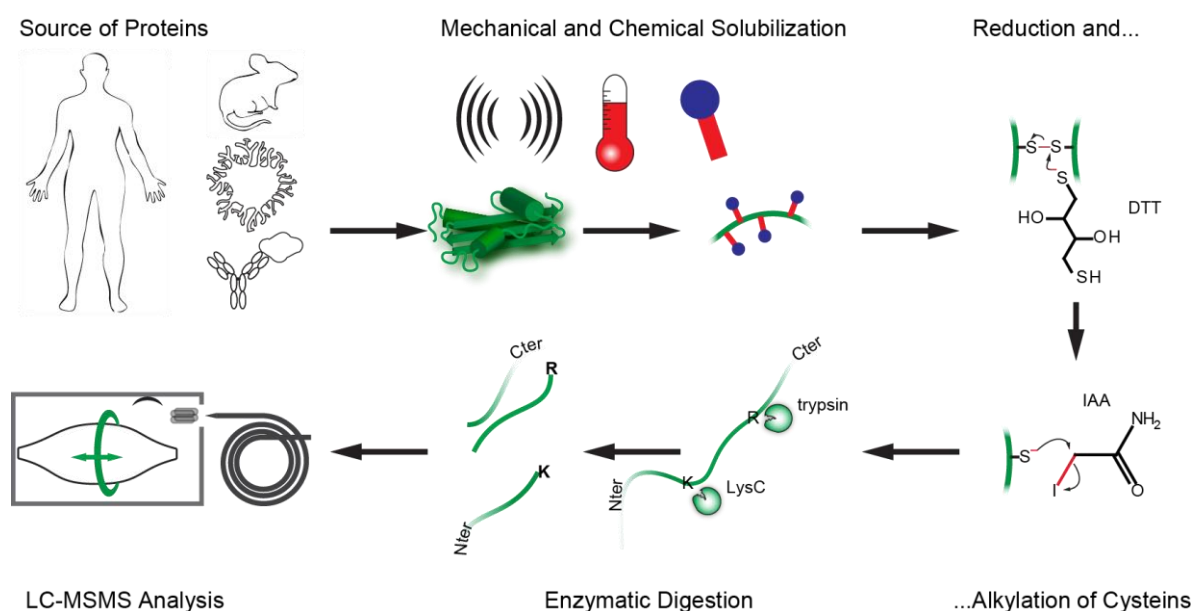

Figure 5 Workflow of an in solution sample preparation: Proteins are isolated or enriched from biological sources. In order to breakdown the cells or tissue environment and to solubilize the proteins, mechanical forces (for instance sonication), increased temperature and chemical agents (detergents or chaotropic agents) are used. Reduction and alkylation may be followed by removal of unwanted buffer components (detergents) and the proteins are enzymatically digested. Resulting peptides are desalted and subjected to the liquid chromatography coupled MS analysis.

Lysis: The purpose of lysing the biological material is to efficiently extract proteins. Depending on the biological question of the study, either the native state is preserved by using physiological buffers with mild detergents such as 1 % NP-40 (for instance for subsequent immunoprecipitation) or all proteins are denatured (complete proteome analysis). The latter approach employs harsh detergence (e.g. 4 % SDS or triton) or chaotropic agents (high molar urea or guanidinium chloride) to denature all proteins and dissolve the lipids. Mechanical breakdown steps such as sonication, bead-milling, rotor stator, blending and heating of the sample often accompany lysis (note however that heating

promotes decomposition of urea and carbamidomethylation of amines). During native lysis, protease inhibitors are often used to avoid premature proteolysis.

Reduction and alkylation: Disulfide bonds and oxidation of cysteins to sulfenic acid, sulfinic acid and cysteic acid expands the number of peptide forms to be considered during analysis. This can be avoided by reducing cysteins with DTT or TCEP and then alkylating them either during or after lysis. Commonly used alkylation agents are 2-iodoacetamide (IAA) and 2-chloroacetamide (CAA). While the kinetic for IAA is faster, CAA is less prone to alkylate amines at lysine and arginine residues. Depending on the proteases used, the alkylation agent should either be removed or quenched.

Sample cleanup: Most detergents need to be removed before proteolytic digestion. In addition, most detergents are incompatible with MS analysis, as they coelute with peptides, ionize very well in electrospray and thereby obscure ionization. Acetone or ethanol precipitation of the proteins or membrane based cleanup (FASP) [142] can be employed if the protein amount is in the low microgram range or higher. For the sub-microgram scale, it is usually advisable to either use MS compatible detergents or chaotropic agents for lysis and to omit the sample cleanup.

Digestion: The most characteristic step of bottom-up shotgun proteomics is the enzymatic digestion of the proteins to peptides. Trypsin is the most common protease in proteomics (see box) but often the lysine specific enzyme LysC is used as well. LysC is active at 8 M urea and has a higher efficiency than trypsin in cleaving C-terminal to lysine. The sequential use of both enzymes allows the digestion at strongly denaturing condition. This facilitates enzymatic breakdown of tightly folded proteins [143]. To maximize sequence coverage (with the penalty of increasing complexity) a separate digest of LysC and trypsin can be performed and peptides can be mixed prior to analysis. Protease activities can be quenched by low pH. In addition to LysC and trypsin, other enzymes can be employed in shotgun proteomics such as chymotrypsin (C-terminal to aromatic residues) or GluC-D (C-terminal to aspartate and glutamate) to increase the peptide coverage of the proteins or to generate peptides with different properties [144]. Before peptides are separated by liquid chromatography (LC) a salt removal and filtering step is performed, based on the principle of solid phase extraction. This also allows fractionation, if desired [145].

The "Gold Standard" Trypsin

Trypsin digestion offers many advantages in shotgun proteomics. (1) It specifically cleaves C-terminal to arginine and lysine (if not followed by proline). Under acidic conditions, this leaves every peptide with at least two positive charges (at their N-terminus and side chain of the terminal residue). The additional charge allows peptide discrimination from usually single charged non-peptide contaminants such as polymers. (2) The stability of trypsin suppresses auto-digestion and allows the proteolysis to take place under mild denaturing condition. (3) The distribution of lysine and arginine in proteomes generates on average ten amino acid long peptides. Peptides with this length have a high chance of uniqueness and are suitable for high resolution analysis in commonly used mass analyzers [8].

2.3.2. Chromatography

After digestion, complex proteome samples contain more than 100,000 unique peptides [132,146]. This peptide complexity cannot be resolved directly in the MS but needs to be separated chromatograpically. This employs the affinity of the analyte (e.g. peptides) to a stationary phase and gradual elution by increasing solubility in a mobile phase. The affinity principle is based either on the coulomb interactions or hydrophobic interaction. Most MS used in shotgun proteomics work in the positive mode, which means that positively charged peptides are analyzed. At low pH peptides have a high affinity to C18-silica in aqueous solutions and can be separated over time by increasing the organic content (e.g. acetonitrile) of the mobile phase.

Chromatography is a critical part of the shotgun proteomics workflow. Reproducible elution profiles and efficient peptide separation have significant impact on the data quality. Downstream bioinformatics analysis attempts to match peptides to each other by their retention times and therefore relies on highly reproducible chromatography (see chapter 2.7). Koch et al. reported an almost linear relation between the number of peptide identifications and the resolving power [147]. The resolving power can be optimized in different ways. **Column length:** Longer columns allow for more interactions between the analytes and the stationary phase (more plates), which in turn increases the resolution [148]. **Column diameter:** A smaller inner diameter increases chromatographic resolution by reducing the number of flow paths (eddy diffusion). However, a smaller inner diameter will inevitably increase the backpressure [148]. **Column filling:** Smaller and more uniform particles as filling material reduce eddy diffusion and the negative effect of mass transfer. **Gradient length:** Longer gradients increase the resolution but can also cause peak broadening and consequently reduction of the ion current due to dilution. Both these negative effects can be compensated with higher flow rates, which however, leads

Resolving Power

A central parameter of chromatography is the resolving power (peak capacity) or simply the number of peaks that can be separated from each other in the gradient. The resolving power is determined by the number of theoretical plates (N_t), which is the number of interfaces between the two phases where exchange of the analytes takes place. The plate number is proportional to the column length (H) and inversely proportional to the *height equivalent to a theoretical plate* (*HETP*), which is given by the Deemeter equation [12]

$$N_t = \frac{H}{HETP}$$

$$HETP = A + \frac{B}{u} + Cu$$

A is the eddy diffusion, which is the variation of analyte channeling speed in the column caused by inhomogeneous column packing. *B* describes the dispersion by random diffusion of the analytes. The slower the flow rate *u* the longer this diffusion can take place, leading to increased peak broadening. *C* is the mass transfer between the stationary phase and the mobile phases. Ideally, this transfer is in equilibrium at any time and the exchange is fast enough to not cause peak broadening by retention of the fraction bound to the stationary phase [21].

to higher backpressure and reduces ionization efficiency [149]. For many years, backpressure was the bottleneck for further improvements of chromatography. With the introduction of new ultrahigh pressure HPLC (up to 1,000 bar) and column heating devices, chromatographic performance has significantly improved [147,150,151].

2.3.3. The Leap of Charges – Electrospray Ionization

In the 1980s, two new soft ionization techniques paved the way for MS based proteomics – electrospray ionization (ESI) and matrix assisted laser desorption ionization (MALDI). ESI is by far the most commonly employed ionization method in proteomics and is based on electrostatic dispersion and desolvation [124,152]. A kilovolt potential is applied between the tip of the chromatographic column and the inlet orifice of the MS. In the positive mode, the column tip supplies the positive charge and tryptic peptides usually carry two or more protons. The solvent of the emitted analytes evaporates, increasing the positive charge density until the electrostatic repulsion releases charged analytes (coulomb explosion at the Rayleigh limit), ultimately transferring single peptide ions into the gas phase. ESI usually results in multiple charged peptide ions with roughly one charge per kDa. Therefore, even large masses can be recorded in a relatively narrow m/z range [153].

> **Mass and Charge**
>
> In a MS individual m/z are determined (see chapter 2.24). From spacing of the isotope pattern the charge and mass can be calculated.

Two principle models are used to explain the generation of gas-phase ions. The ion evaporation model suggests that the increase in charge density at the surface cause jet-like ion desorption. The second model (charge residue model) postulates evaporation and fission cycles of the droplets until single desolvated analytes remain [154]. These processes are not mutually exclusive. A growing body of evidence suggests that large molecules are released according to the charge residue model whereas smaller ions are released into gas-phase via ion evaporation [155].
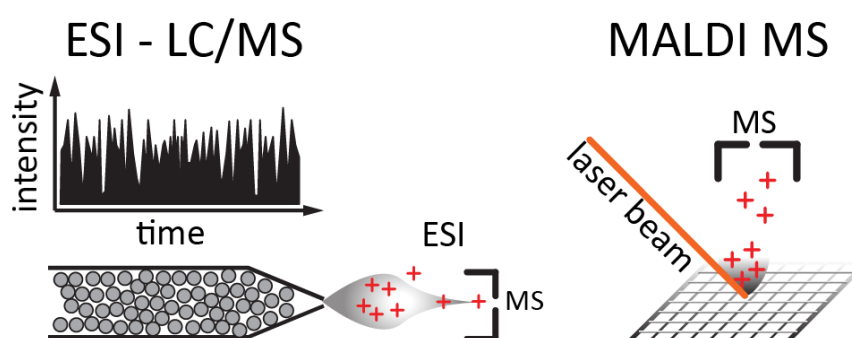

Figure 7 Gentle ionization techniques MALDI and ESI: In ESI ions are transferred into the gas phase by desolvation from electrically charged droplets. ESI can easily be coupled to liquid chromatography, which is very useful for the analysis of complex samples. MALDI is the other gentle ionization technique. It uses a pulsed laser to evaporate analytes incorporated into a matrix. MALDI primarily generates singly charged peptides. Compared to ESI, MALDI is more robust with respect to buffer conditions [10].

2.4.   The Mass Spectrometer

2.4.1.   Mass Analyzer

In proteomics different types of mass analyzers are employed (Figure 8). Two groups of mass analyzer can be distinguished. The trap based mass analyzers include 3D and 2D ion traps (linear ion trap quadrupole, LTQ), Fourier transform ion cyclotrons (FTICR) and the Orbitrap analyzer. The beam based mass analyzer are made up of the triple quadrupole and time of flight (TOF) instruments. In proteomic practice, hybrid mass spectrometer contain elements of each group. They consist of different segments for ion guidance, selection, fragmentation and detection, combining the strength of the individual technologies. The ion optics that enable ion guidance and the mass analyzers themselves rely on a set of physical principles that govern ion motion. The following parameters are used to describe the performance of mass analyzer.

Mass resolution: High mass resolution allows to distinguish two peaks with very similar m/z. This is very important for both the mass accuracy and the quantification as it is necessary to sufficiently resolve the contributing components [156]. The resolution is defined as the ratio between the m/z of a peak and the delta m/z at the full width half maximum (FWHM) and is therefore a dimensionless quantity [157]. It varies between mass analyzers and is m/z dependent for some of them. In proteomics 'high resolution' usually refers to > 10,000 and has now become standard in the field [158]. **Mass accuracy**: The molecular mass is the sum of the atomic masses. The exact mass measured by MS for each isotope is different from the chemical mass where the abundance weighted average isotope masses are used. The mass accuracy refers to the deviation between the theoretical mass and the experimentally determined mass and is expressed as the mass error in parts per million or billion (ppm, ppb). The mass accuracy depends on many parameters such as the resolution and signal to noise ratio [156]. Systematic mass errors can be corrected by internal and external calibrations [159]. In proteomics, unambiguously identified peptides can serve as internal calibration points [160]. The **dynamic range** of an instrument is an important parameter as it indicates how well the mass analyzer detects low abundant ions in a mixture with very high abundant ions. Especially in very complex biological samples, the abundance differences of proteins is many orders of magnitude. **Scan speed**: This parameter describes how fast a certain m/z range can be monitored and for many instrument types, is roughly inversely correlated with the resolution. **Sensitivity**: The detection limit of a mass analyzer depends on the principle of detection and in the proteomic context, also on the sample complexity and therefore dynamic range.
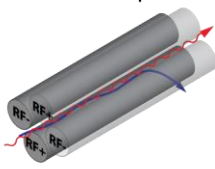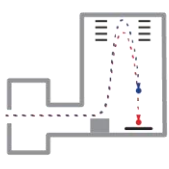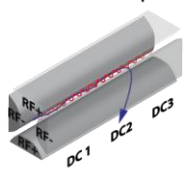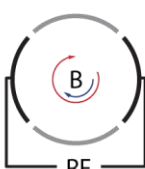
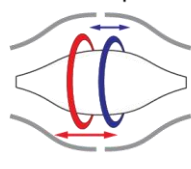|	Beam type		Trap type		
	Quadrupole	TOF	linear ion trap	FTICR	Orbitrap
detection	stable trajectory	velocity and time	stable trapping	axial rotation	axial osciliation
max. resolution	500-2,000	60,000	1,000-5,000	>10,000,000	1,000,000
mass accuracy	ppm	low ppm	ppm	high ppb	high ppb
dynamic range	4-5	3	2-3	3	3
scan speed	high	very high	medium	low	high
sensitivity	very high	high	very high	high	high

Figure 6 Commonly used Mass Analyzers: Mass analyzer can be classified into beam type and trap type analyzers. TOF, FTICR and Orbitrap analyzers are usually referred to as high-resolution analyzers. The underlying technologies have various advantages for mass spectrometric applications (adapted [161,162]). Different m/z are discriminated via their trajectory. This is achieved by changing electrical fields (radio frequency current (RF) and direct current (DC)) for Quadrupole and Linear Ion Trap, electrostatic fields for Orbitraps and TOFs or magnetic fields (B) and RF for the FTICR. A more detailed description of these technologies is provided in the following chapters. Parameters are provide for general guidance only.

2.4.2. Beam Type Mass Analyzers

One of the oldest beam type mass analyzer combines the mass selection capacity of a quadrupole mass filter with an ion detection unit such as electron multiplier. The quadrupole consists of four parallel rods with two opposing rods having a positive and the other two rods having a negative potential on which a time dependent radio frequency potential (RF) is superimposed. The motion of ions in this quadrupole field is governed by the *Mathieu equation,* a second order differential equation derived from the Lorentz Force and Newton's Law for this electrical field. The *Mathieu equation* features two parameters **a** and **q,** characterizing the amplitude of direct current (DC) and RF voltage, respectively. Plotting **a** against **q** provides a stability diagram for m/z. While unselective ion guidance can be achieved with $a = 0$ (RF only mode), $a \neq 0$ is used to stabilize a specific m/z range.

The second beam type instrument is the time of flight mass analyzer. Unlike the quadrupole, which separates different m/z by path-stability, the TOF separates m/z by velocities in a field-free drift tube. Ions are accelerated in an electric field providing ions of the same charge the same kinetic energy. Because this kinetic energy is determined by mass and velocity, the arrival times of ions increases with mass. Generally, the longer the drift field and the faster the detector, the better the resolution that can be achieved in a TOF. In addition, the spatial and energetic drift can be reduced by electrostatic mirrors pushing state of the art TOFs to medium or high resolution performance (>30,000) [163].

Ion Motion

From the ionization source (ESI or MALDI) charged peptides are introduced into the MS. Two physical principles govern ion motion. The Lorentz force and Newton's second law. In an electric and / or magnetic field the force $\vec{F}$ a charged particle experiences is determined by

$$\vec{F} = z(\vec{E} + \vec{v} \times \vec{B})$$

Where z is the electric charge of the particle that experiences the electric field $\vec{E}$ as well as the vector-cross product of the velocity $\vec{v}$ and the magnetic field $\vec{B}$. The acceleration $\vec{a}$ is given by

$$\vec{F} = m * \vec{a}$$

and the trajectory of ions therefore depends on the mass-over-charge ratio m/z.

$$\vec{a} = \frac{(\vec{E} + \vec{v} \times \vec{B})}{m/z}$$

The m/z dependency allows to selectively manipulate trajectories of ion species applying specific electric or magnetic field combinations.

2.4.3.   Trap Type Mass Analyzer

In contrast to beam type analyzers, ion traps can store and accumulate ions over a period of time. A 3D ion trap conceived in 1953 by Paul and Steinwedel was a basis for MS based proteomics for many years. The 3D ion trap consists of three electrodes, a hyperbolic ring electrode and two end-cap electrodes. Ions entering this trap are slowed down by collisional cooling with an inert gas and are confined in a 3-dimensional quadrupole field created by a time varying potential applied to the end caps (RF) [164-166]. The 3D ion trap was superseded by the 2D ion trap, which stores ions in two dimensions along the z-axis of a quadrupole. The 2D ion trap is very similar to the quadrupole mass filter but it contains the ions by DC potentials at its entrance and exit (Figure 8). The different geometry of a 2D trap offers a longer trapping path, which increases ion-storage capacity more than tenfold. Decreased space charge effects in 2D ion traps lead to higher resolution and mass accuracy [167-169].

Typically, a 2D ion trap is employed for ion accumulation and storage in hybrid mass spectrometers, whereas its use as stand alone mass analyzer is compromised by relatively low resolution and mass accuracy. The breakthrough for high-resolution mass analyzers in proteomics was achieved with the introduction of the hybrid linear ion trap – FTICR. In the LTQ FT the linear ion trap is coupled to an ion cyclotron resonance (ICR) cell, which confines ions in a static electric field along the axial dimension and forces ions on an orbital trajectory with a strong magnetic field (typically 7 T or more). From this periodic rotation, the m/z value is determined. Although the FTICR features very high resolution, it is very bulky, expensive and has a relatively low ion storage capacity. In proteomics, most of these limitations were addressed by exchanging the ICR with the introduction of hybrid Orbitrap mass analyzers.

The Orbitrap Analyzer

In 2000 a novel mass analyzer, the Orbitrap, was introduced by Makarov [170,171]. This ion trap is derived from a simple design by Kingdon [172] and features dynamic trapping in an electrostatic field (not using any RF field in contrast to the quadrupoles or 3D or 2D ion traps) around a cone shaped electrode (Figure 9). The electrostatic field generates a quadro-logarithmic potential distribution and is composed of a quadrupole field generated by the ion trap outer barrel-like electrodes and the field of the spindle-like inner electrode [171]. The axial field strength is zero at the spindle equator and gradually increases towards the outside. Ions are injected off-center and circulate around the central

electrode on stable trajectories and oscillate harmonically along the z-axis [13]. The frequency of only the z-axis oscillation is completely independent of initial variables such as initial ion velocity and position and it is inversely proportional to $\sqrt{m/z}$ [171]. Injection within a time interval that is small in relation to the oscillation frequency forces ions of the same m/z to oscillate along the z-axis in phase for hundreds of thousands of periods shaping a thin ion ring orthogonal to z after radial dephasing (20-100 μs) [13,159,173].

Radial and rotational motions induced image currents cancel each other due to the symmetric shape of the ion ring [173]. Meanwhile the axial oscillation of the ion ring from one-half of the outer electrode to the other induces opposing currents in the outer electrodes, which are composed of two electrically separated halves. In contrast to 2D and 3D ion traps, which detect ions sequentially, the Orbitrap therefore records an alternating image current induced in the outer electrode by all trapped ions over a certain time (transient time) [174].

Detection Limit of the Orbitrap

The image current detection of the Orbitrap initially had a noise-band equivalent to 20 ions at 1 s transients, which is less sensitive than the theoretical single ion detection capability of electron multipliers [13]. However, a proof of principle study published in 2009 showed single ion detection capacity of the Orbitrap for highly charged and purified ions [15].

To deconvolute the information of all m/z in the Orbitrap, a Fourier transformation (FT) is applied. FT is a set of mathematical operations to approximate complex periodic functions (= signals) as an overlay of different sinusoid functions. The periodic nature of the monitored image current dictates the dependence of the resolution on the number of oscillations along the z-axis and therefore on m/z [157,171,174]. Compared to FTICR, the resolution in an Orbitrap analyzer decays with $\sqrt{m/z}$ rather than with m/z [156].

The maximum resolution in an Orbitrap is also affected by other factors. First, the time of the transient is limited by dephasing of the z-axis oscillation of the individual ions in the Orbitrap. Dephasing is promoted by space charge effects, imperfection in the electrostatic field and collision with other molecules. The latter makes dephasing dependent on the size of the molecule [159]. All these factors contribute to axial broadening of the ion ring, which ultimately cancels out the image current [173]. In current systems, an image current in the Orbitrap becomes indistinguishable from noise after one second [156].

During a transient newly arriving ions cannot be analyzed. Therefore, long transients increase the chance to miss ions in complex mixtures. Because the resolution in an Orbitrap is inversely proportional to the speed of the instrument, one might choose a lower resolution for complex mixtures to detect and sequence more peptides eluting in a narrow time window [15].

Over the years Orbitrap performance has significantly improved by the development of enhanced Fourier transformation (eFT) and by implementing a smaller, high-field Orbitrap analyzer [175]. The eFT takes phase information into account, which provides a roughly two fold increase in resolution for the same transient length [176]. The smaller dimension and stronger electrostatic field of the high field Orbitrap analyzer increase the oscillation frequencies of ions and allows to boost either the speed or resolution roughly by a factor two.
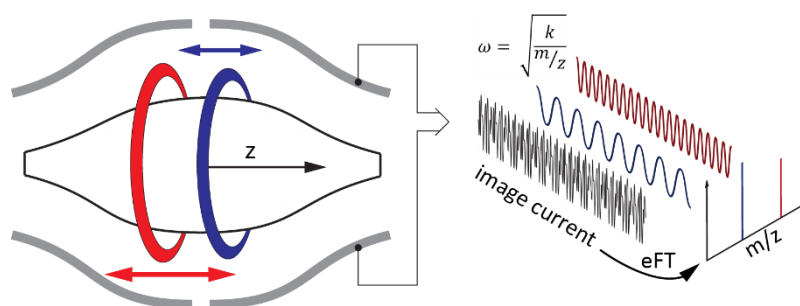

Figure 7 Orbitrap mass analyzer: This analyzer records the axial oscillation of trapped ions. The number of oscillations is directly proportional to the resolution. This frequency depends on the instruments parameters (k) and decreases with the square root of m/z. The recorded image current contains the overlay induced image currents of all m/z species in the Orbitrap. Signals for individual m/z are extracted via enhanced Fourier Transformation [156].

2.4.4.   Ion Fragmentation

Another central feature of modern hybrid MS is their ion fragmentation capability. Fragmentation of ions and determination of the fragment masses provides an additional layer of information beside the mass of the precursor ion, which is extracted from the mass spectra or MS^1 spectra. The identification of proteins via their proteolytic peptides can be based on the MS^1 (parent-ion mass) information alone, which is called "peptide fingerprinting". However, this approach is accurate only for isolated proteins or simple protein mixtures and clearly not suitable for analyzing complex protein mixtures. This is because many peptides of different sequence have very similar masses and peptides with the same amino acid composition but different primary structure are indistinguishable even in principle by this approach. To increase the information on the sequence of individual peptides, mass spectrometric fragmentation is employed (this is termed MS/MS or MS^2). Peptides of interest are isolated within a narrow window around the desired m/z, are subjected to fragmentation and these fragments are measured in a subsequent mass spectrum [177,178].
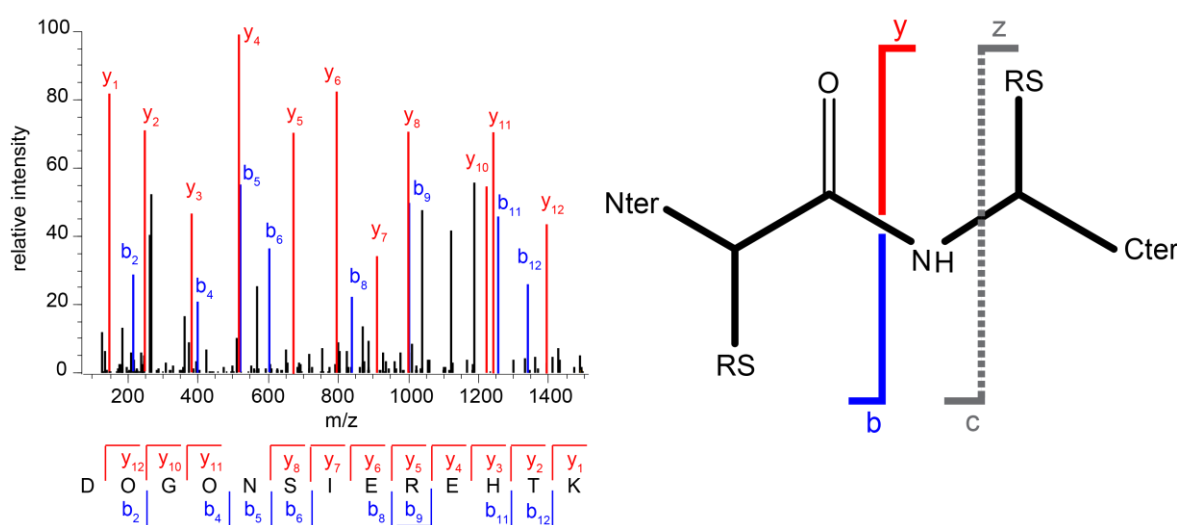
Figure 8 Fragmentation spectrum of a peptide: A typical HCD fragmentation spectrum shows the b-and y-ions series of the peptide. While the y-ion series is almost complete, only some b-ions were identified for this peptide. Higher coverage of the y-ions is characteristic for HCD. The cleaved bonds for y-ions (red) b-ions (blue) as well as for z- and c-ion (grey) which are generated in ECD / ETD fragmentation are shown in the lower panel.

A common fragmentation principle is to induce dissociation of the peptides by collision with an inert gas such as He or N_2. In this process, the kinetic energy is partially converted into internal energy, which eventually breaks the chemical bonds (Figure 10). The collisional energy required for efficient fragmentation depends on peptide mass and charge state. Two commonly used fragmentation methods, higher energy C-trap dissociation (HCD) and collision-induced dissociation (CID), preferentially fragment peptides at the amide bond to series of b-and y-ions, the N- or C-terminus, respectively. If complete, each of those would independently provide the complete sequence information [179].

Although HCD and ion trap CID are related, they have conceptual differences [180]. When CID is performed in an ion trap, the parent ions are resonance-excited and fragmented. Due to activation of only a specific m/z, further fragmentation of the fragments is rare. Fragmentation in the ion trap is fast and efficient but can suffer from a predominance of only the lowest energy fragment as well as from its low-mass cutoff [181,182]. This low-mass cutoff is due to unstable trajectories of fragments with less than 1/3 of the precursor m/z [183]. In contrast, HCD, which is very similar to triple quadrupole CID, takes place in an RF multipole, which does not activate ions via resonance excitation but fragments via a beam type collision, allowing even very small fragments to be observed. HCD also often entails further fragmentation of the fragment ions, which can complicate spectra but also

help provide additional sequence information. For these reasons, the chemically less stable b-ions are often further fragmented, leading to prominent y-ion series in HCD [177,180]. While ion trap CID fragmentation spectra are usually recorded at low resolution and low mass accuracy, HCD usually features high resolution and high mass accuracy analysis of the fragment ions in the Orbitrap [177,179].

Apart from HCD and CID, several other fragmentation principles are used in mass spectrometry. A common orthogonal approach is the activation of peptides with electrons. Both electron capture dissociation (ECD), which is employed in FTICR instruments and electron transfer dissociation (ETD), which is used with ion traps, are based on this principle. In ECD peptides are directly bombarded with electrons while radical anions are used as electron carrier in ETD. Both methods break the N-C bond and generate series of c-and z-ions [184,185]. ETD and ECD have advantages in the analysis of intact protein and of peptides carrying PTMs where it is desirable to avoid fragmentation of weak bonds [186]. The orthogonality of ETD / ECD compared to CID / HCD can be very useful to increase the information about the parent ion, a critical aspect in some proteomics approaches such as *de novo* sequencing of peptides. HCD and ETD can also be combined at the same time for extensive sequence information [187].

2.5. The Q Exactive High Field

The analysis of complex proteomes containing thousands of peptides after tryptic digest greatly benefits from the highest performance hybrid mass spectrometers. The latest generation of hybrid Orbitrap MS is called Q Exactive and couples a quadrupole mass filter, rather than a linear ion trap, with an Orbitrap analyzer [188]. This instrument features high resolution in both MS^1 and MS^2 modes as well as extremely fast MS^2 performance. Its latest version, the Q Exactive high field (HF) consists of an advanced quadrupole and a high field Orbitrap analyzer making it especially well suited to complex proteome analysis and is discussed and evaluated in detail in *article 6*. Its mode of operation is briefly described with reference to Figure 11.

Peptides ions from an electrospray source enter via the orifice and the **S-lens**, which is composed of a series of flat-ring electrodes with an alternating RF of opposite phase. It efficiently transfers ions from the atmospheric pressure of the ESI to the low pressure in the MS. The spacing progressively increases to confine ions in an increasing RF field [189]. The optimal RF for focusing depends on the m/z and is adjusted for the individual precursor or precursor range. The S-lens increases the ion current by several fold compared to the previous set-up [177]. Many of the background ions entering the MS are not of interest for the analysis but can cause problems by contaminating the ion path. Removing those as early as possible therefore ensures long-term performance and increases

cleaning intervals. For this purpose, the Q Exactive HF features two early filters. The **injection flatapole** has low-resolution ion selection capability and is located near the front end of the instrument. It allows removal of unwanted ions outside the m/z range of interest prior to the more sensitive parts of the MS. The **bent flatapole** excludes neutral molecules and droplets from the ion beam, because they cannot follow the bent trajectory. While older Orbitrap hybrid MS were equipped with a linear ion trap [13,175,177,189] the Q Exactive HF instrument features a **quadrupole mass filter** with excellent ion transmission and mass selection capacity. It filters ions within a specified m/z window with nearly perfect isolation shape and passes them into the C-trap. The **C-trap** is situated between the **Orbitrap analyzer** and the **HCD cell**. The C-trap is a curved RF only quadrupole. It can accumulate and store ions prior to injecting them into the Orbitrap. These ions can either come directly from the ion source (MS[1]) or they can be fragments produced in the HCD cell, which are then injected into the Orbitrap (MS[2]). For accumulation, ions entering the C-trap are confined by the trap electrode, cooled by collisions with nitrogen and come to rest along the curved axis. Injection of the focused ion package into the Orbitrap is accomplished by applying a very sharp DC voltage spike. More than one million ions can be stored in the C-trap before space charge effects distort energy and spatial distributions, preventing further accumulation. To prevent over or under loading of the C-trap, the ion accumulation is regulated by automatic gain control (AGC). This AGC value is usually set between 1 and 3 million for MS[1] and about 50,000 for MS/MS. Compared to 2D traps, the ion storage capacity is more than an order of magnitude higher [13,190].
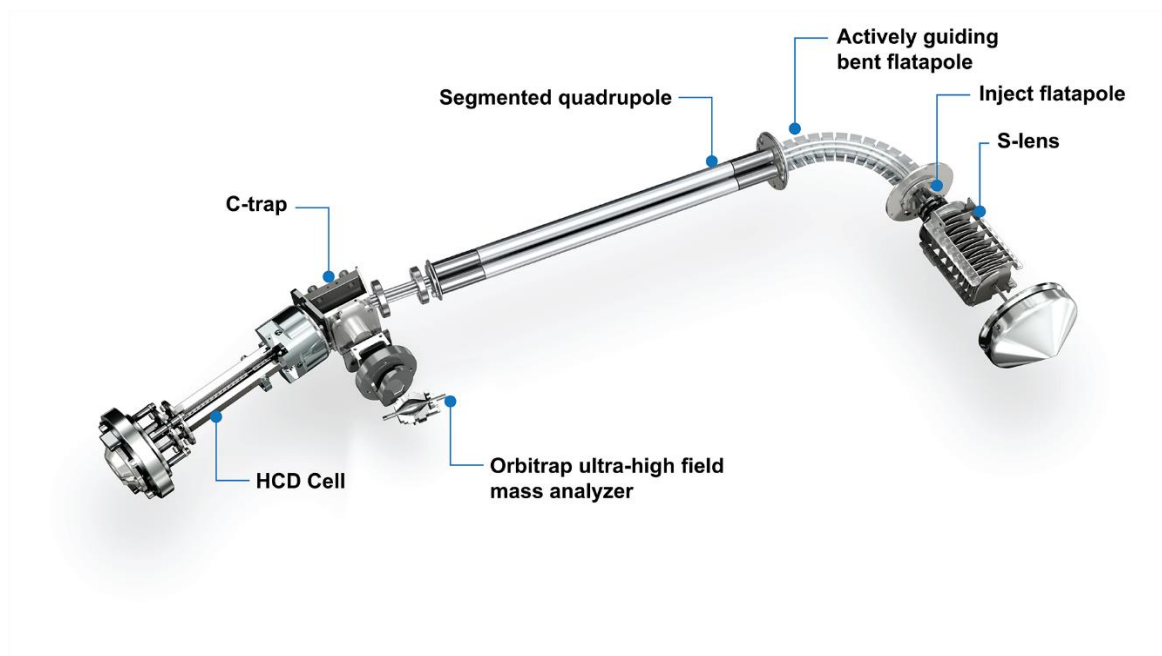

Figure 9 Q Exactive HF: Ions are injected and focused in the RF lens (S-lens). The injection flatapole is a low resolution mass filter. Neutral particles are removed from the ion beam in the bent flatapole. The segmented quadrupole acts as mass filter for ions to be subjected to direct mass analysis via the C-trap or via intermediate HCD fragmentation [5].

2.6.   Acquisition Strategies

In shotgun proteomics, data dependent acquisition (DDA) strategies have generally been used. The most used method is called topN because eluting peptides are measured in a survey scan (MS^1 or full scan) followed by selection and fragmentation of the top N most abundant peptides that have not been fragmented before. On our instrumentation in particular, the survey scan usually covers a m/z range between 300 to 1650 Th at a resolution of 60,000. Higher resolution (120,000) can improve the efficiency of match between runs algorithms (see Chapter 2.7) but will reduce the speed of the instrument. After each MS^1, the 5 to 20 most intense features with a charge state higher than one are sequentially subjected to fragmentation and the fragment masses are recorded in separate MS^2 spectra. In the Q Exactive, fragmentation is performed in the HCD cell and spectra are obtained at a high resolution (for MS^2 scans) of 15,000 in the Orbitrap as opposed resolution less than 1000 in older instruments that perform CID fragmentation and mass analysis in the linear ion trap [190]. The two approaches are termed High-High and High-Low strategy, respectively. In the Q Exactive HF, an acquisition cycle, consisting of one MS^1 and 15 MS^2 scans, takes roughly one second. This is fast compared to the several second that it takes peptides to elute from the chromatographic column. To prevent resequencing of the same peptide, precursors with the same mass are excluded from selection for 20 to 40 seconds. If the samples are very complex, under-sampling can occur. This gives rise to the 'missing value problem' in DDA, where some peptides are sequenced in one but not the other identical sample. With this background, data independent acquisition (DIA) strategies like SWATH MS have been developed, in which the mass spectrometer selects a precursor range of about 10 to 25 m/z units and cycles through the mass range [191,192]. This has the advantage that in principle all precursors are sequenced, however, so far at the cost of diminished dynamic range. The complex nature of the resulting MS^2 spectra for each isolation window and elution time requires a peptide fragmentation library to infer peptide identities. SWATH MS is not hypothesis driven and allows to screen the spectra for other information post acquisition providing a suitable library. However, data interpretation is much more complicated. For SWATH MS quadrupole-TOF instruments are generally used because they are fast enough to sample the mass range in a time window smaller than the average time for peptide elution

A third mode of bottom-up proteomics is used in a 'hypothesis driven' or 'targeted' mode and is called single or multi reaction monitoring (SRM or MRM). Pre-defined precursor ions are subjected to fragmentation to monitor specific precursor / fragment ion pairs without scanning at the MS^1 or MS^2 level. Usually triple quadrupole instruments have been employed in targeted proteomics. In this setup, the first quadrupole serves as precursor ion selector, the second as the collision cell and the third as an ion selector, which passes a predefined fragmentation product to the detector. This

approach has a very high sensitivity but relies on prior knowledge on the exact mass and retention time and is so far limited to the detection of approximately 100 proteins per sample [161,193]. Furthermore, specificity of detection can be a challenge in SRM or MRM. This is alleviated in 'parallel reaction monitoring', in which the precursor ion is targeted but a full scan of the fragments at high resolution is taken [194].

2.7.   Protein Identification

The interpretation of the spectra recorded in the MS is a complex procedure that is nowadays performed in an automated manner. Peptide identities often cannot be inferred unambiguously from their accurate masses as discussed in chapter 2.4.4. Ideally, each MS^2 provides complete fragment ion series for each peptide, which would in principle allow to determine the exact sequence based on this information (*de novo* sequencing). In practice however, *de novo* sequencing is not feasible in routine workflows. MS^2 spectra are contaminated with co-isolated precursors, contain internal and side-chain fragments, and are often incomplete. To infer the sequence identity from the MS^2 and MS^1 spectra from a s sequence data base, search engines like Andromeda or Mascot are employed [160,195]. These search engines perform an *in silico* digest of all possible proteins in the proteomes studied and calculate theoretical MS^2 spectra. Constrained by the MS^1 masses, spectra recorded in the MS^2 are matched to these theoretical spectra. They are evaluated using a probability based scoring model. MaxQuant, for instance, calculates posterior error probability (PEP) for each peptide based on the identification score and length [196]. The quality of this match strongly benefits from high resolution data with high mass accuracy as this decreases the search space. Apart from employing high performance instrumentation and workflows, the mass accuracy can be improved *in silico* by averaging masses obtained in consecutive surveys scans as well as by performing an *in silico* mass recalibration based on peptide charge pairs. These algorithms in combination with Orbitrap instrumentation achieve average absolute mass accuracies below 400 ppb in the MS^1 [197].

> **Matching between runs**
>
> The partially stochastic nature of the DDA leads to missing MS^2 identification. A robust workflow allows to transfer identification from one sample to another by aligning the retention times of MS^1 features. This algorithm is called 'match between runs' in the MaxQuant software suite.

A proteome dataset can contain hundreds of thousands of features that are scanned against the database as described above. For such large datasets, it is essential to evaluate the likelihood to obtain a good score for an identification purely by chance. To estimate the likelihood of those false positive hits, a 'target decoy' false discovery rate estimation is performed. The standard MaxQuant procedure

is to combine the protein reference database with a library containing the reversed amino acid sequences. In addition, in the reverse data base each lysine and arginine (for tryptic digests) is swapped with the preceding amino acid. This avoids having exactly the same mass for the reverse peptide as for the respective forward peptide while preserving the general context of each amino acid [197]. Spectra are matched to this combined database and the number of hits in the reversed sequences and the corresponding score can be used to define the cutoff and ensure a desired false discovery rate (FDR) – usually 1 % [196,198].

All peptides are assembled into individual proteins. Proteins that cannot be discriminated by unique peptides are combined in proteins groups. Some peptides are ambiguous regarding the corresponding protein. According to *lex parsimonies* these peptides can be attributed to the protein group that already has the highest number of peptides. Assembled proteins are also FDR controlled based on the ranked protein PEP, which is the product of the respective peptides PEPs [160,196]. The protein FDR usually dominates the peptide FDR, so that proteins that are confidently identified have peptides with FDRs much less than 1 %. The importance of the protein FDR concept was illustrated by two publications that claimed to provide the most comprehensive draft of the human proteins so far, covering roughly 18,000 proteins [199,200]. However, these did not employ conventional protein FDR calculations, raising the possibility that the overall number of identified proteins is boosted at the expense of data quality. This is suggested by the surprisingly high number of olfactory receptors, that these studies reported in various tissues in which they are not thought to be expressed [201].

2.8.   Protein Quantification

The identification of peptide and protein species is only the first step in the process of gaining insight into the biological system studied by proteomics. Proteins in the cell cover an abundance range of more than six orders of magnitude and interpretation of biological phenomena require quantitative information. *Article 1,* for instance, illustrates how quantitative information is essential to rigorously evaluate the similarity between cellular systems. Two levels of quantitative information can be distinguished: First, 'relative quantification' measures the difference between the same proteins in two or more samples. Second, 'absolute quantification' determines the absolute amount of the proteins of interest in copy numbers per cell or as a concentration. This is currently done by correcting the measured peptide signals identifying one protein for physical difference of individual proteins (for instance number of possible peptides). This type of absolute quantification provides information about the dynamic range of the measured proteome and ranks every measured protein in the proteome by expression level. Over the years, various quantification strategies have been developed that differ

significantly with respect to the method, accuracy and the quantification level. One discriminating feature is whether they are based on isotope labeling or on label-free approaches.

2.8.1.   Label-based Quantification

Labeling strategies are often applied to derive highly accurate relative quantitative information. Labeling with an isotope tag allows combining samples and processing them together, thereby reducing the variation introduced by separate sample preparation. These isotopes do not significantly change the chemical property (chromatography or ionizability) and they can be distinguished in the MS [202]. Approaches can be divided into metabolic and chemical labeling. Metabolic labeling such as stable amino acid labeling in cell culture (SILAC) allows combination of samples right after the biological experiment, reducing variances due to sample preparation to a minimum [203]. In SILAC, one set of cells incorporates a heavy isotope labeled version of an essential amino acid, whereas the other set of cells incorporates the normal amino acid. Heavy arginine and heavy lysine are commonly used allowing nearly all resulting tryptic peptides to be distinguished by the respective m/z offset in the MS^1 spectrum and their corresponding intensity differences to be used for very robust and accurate quantification. SILAC labeling is also possible in whole organisms such as mouse and the technology allows monitoring protein turnover (pulsed SILAC) [204,205]. Although unmatched in accuracy SILAC, has some limitations. One caveat of SILAC is that the MS^1 spectra have twice the complexity. Furthermore, SILAC initially only allowed the comparison of two conditions and could not be used to make quantitative comparisons in every biological system, as it required the incorporation of at least one of the heavy amino acids without metabolic conversion. The introductions of super SILAC – which uses mixtures of SILAC labeled cell lines as a spike-in reference – extended the use of SILAC to various biological systems that cannot be metabolically labeled [206].

Chemical labeling can be performed independently from the source of the material and allows to combine samples either before or after enzymatic digestion. Different chemical labeling strategies have been developed. Similar to SILAC, iCAT and dimethyl labels lead to peptide pairs in the MS^1 spectrum, however, iCAT only labels cysteins at the protein level, leading to decreased sample complexity [207,208]. The label further contains an affinity handle (such as biotin), which allows enrichment. Unlike iCAT were quantification is solely based on the cysteine containing peptides, dimethyl labeling allows quantification on all peptides. It employs isotopically labeled formaldehyde to react via reductive amination with ε-aminogroup of lysine and the N-terminus. In contrast to iCAT and dimethyl labeling, isobaric labeling strategies such as TMT and iTRAQ use the MS^2 level to

distinguish and quantify the different peptides [209,210]. Isobaric tags have the same chemical properties and masses in the MS^1, but result in distinct fragment ions in the MS^2. This strategy has the advantage of not increasing the sample complexity on the MS^1 level and allowing higher multiplexing, with 10-plex kits now commercially available. However, unless special measures are taken, MS^2 quantification suffers from ratio compression due to coeluting peptides that contribute to the intensity of the tag-ion intensity [202]. In addition, the reporter ions have relatively low masses (<150 Da), which tend to be unstable in conventional ion trap (1/3 rule) CID fragmentations. This stability problem however, was solved with the development of new technologies such as pulsed Q dissociation and ETD and is in any case not an issue when using HCD [202].

2.8.2.   Label-free Quantification

Label-free comparison of proteomes has several practical and conceptual advantages, such as being applicable to any sample, avoiding increase in proteome complexity, being freely scalable to any number of comparisons and increasing accuracy for very high ratios. However, it entails completely separate processing and measurement of the samples, which reduces the accuracy and robustness of the quantitative information unless the workflow is very well controlled.

Early label-free quantification methods were based on the correlation between the number of spectra recorded for a protein and its abundance. In spectral counting the abundance of proteins is estimated by the number of MS^2 spectra recorded for each protein [211]. However, this quantification method requires high numbers of fragmentation spectra for each protein to be at least moderately accurate. Alternatively, the number of identified unique peptides can be used as an estimator of protein abundance. The protein abundance index (PAI) normalizes the number of identified peptides by the number of theoretically observable peptides, providing a rough estimation of the abundance. Furthermore, 10^{PAI}-1 empirically turns out to be roughly proportional to absolute protein abundance [212]. APEX is another algorithm to estimate the abundance of proteins based on the number of observed peptides. In contrast to PAI, APEX uses a machine-learning algorithm to predict the observable peptides to estimate absolute abundances [213]. Although relatively easy to implement, neither spectral counting nor the number of identified peptides comes close to the accuracy achieved in labeling approaches such as SILAC.

More recently developed label-free strategies make use of high resolution data and employ the MS^1 ion intensities of all the identified peptides (extracted ion current or XIC) to extract quantitative protein information. The MS^1 intensity is directly proportional – within the linear dynamic range of the instrument – with the number of ions. The MaxQuant MaxLFQ algorithms calculate ratios of

normalized peptide intensity and uses those to obtain relative or absolute quantitative information from the MS data. These sophisticated algorithms, in combination with robust and simple workflows as well as high performance MS, often allow to quantify protein changes within a fold-change of two, which is almost in the range of SILAC accuracy [214].

2.8.3. Copy Number Estimation

The quantification strategies mentioned above provide limited information about the copy number of proteins. Apart from the number of cells, this level of information requires normalizing all proteins with respect to each other and usually requires a reference to a standard with known amount. Spike-in of isotopically labeled standard peptides, whose molarity has been determined for example by amino acid analysis can be a very accurate way to determine protein copy numbers of the respective proteins. The Protein Epitope Signature Tag (PrEST) strategy makes use of the Human Protein Atlas Project where parts of all human proteins had been produced for generating antibodies [215]. PrESTs consist of unique sequence stretches of the proteins of interest, fused to a common Protein A derived solubility tag that has no equivalence in the proteome. PrEST can be isotopically labeled and quantitative information can be derived from the ratio of the endogenous protein to the PrEST in the MS^1 spectrum. The exact amount of the spiked in PrEST can be determined via the solubility tag, which is repurposed as a quantification tag. In comparison to related approaches like AQUA or QconCAT, PrESTs have the advantage of a common reference tag (solubility tag) and a preexisting library [216-219]. Even more importantly, PrEST correct for the unknown and varying efficiency of digestion of proteins to peptides. Other approaches rely on normalized summed peptide intensity in combination with a spike-in standard (iBAQ [220]) or the summed intensity of the three most abundant peptides normalized with an internal standard (Top3 [221]). Finally, the total protein abundance (TPA) approach does not require any spike-in standard and allows estimating copy numbers purely based on XIC quantitative information, the molecular mass and the estimate protein amount per cell [222]. In summary, copy number estimation is possible with different levels of accuracies depending on the method used. Highest accuracy is achieved by using spike-in standards. Normalizing all proteins with respect to each other allows to interpolate the abundance of proteins without a corresponding reference, however with reduced accuracy.
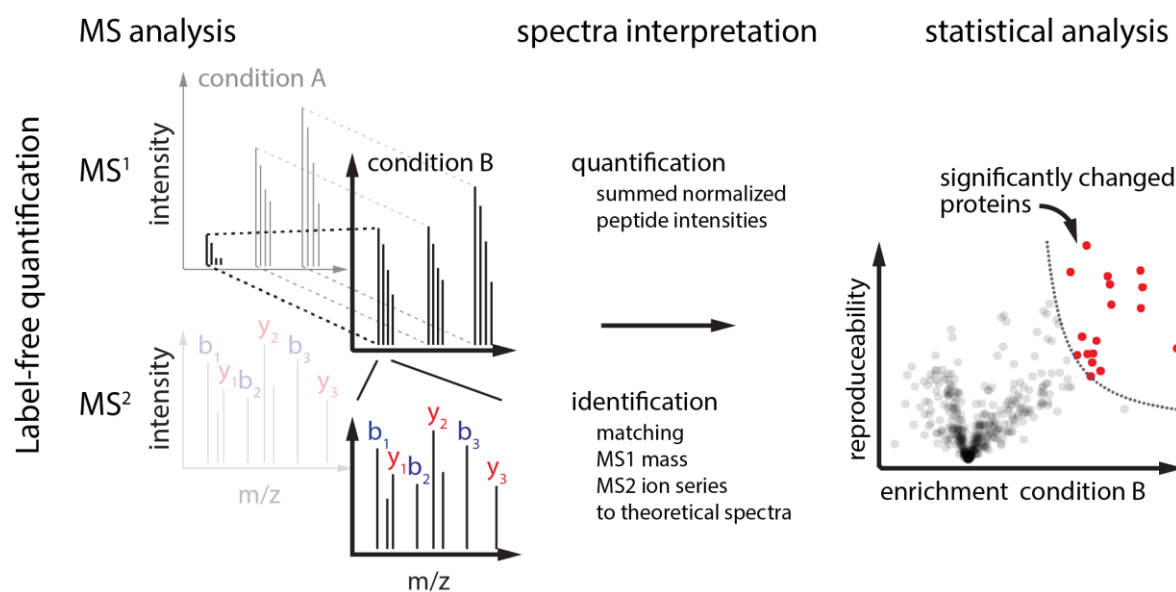
Figure 10 MS feature interpretation: MaxQuant fits a Gaussian curve to the three central data points of each MS[1] peak and calculates the intensity weighted mass deviation and assembles 3D peaks of each eluting feature visible in the spectra. High mass accuracy significantly reduces the search space and improves data quality [223]. For quantification, the apex intensity of the 3D peak is employed. In contrast to the integral of the MS[1] peak, the apex is more robust towards background peaks and non-symmetrical peak shapes. On the basis of the MS[1] features and the corresponding MS[2] spectra, peptides are identified, assembled to quantified proteins and quantitative difference are subjected to statistical testing.

2.9.   Optimized Experimental Designs

Complex systems like a MS have many adjustable parameters that may or may not significantly affect their performance. For instance, as discussed in chapter 2.4.3 high mass resolution is a precondition of high mass accuracy but will inevitably reduce the sequencing speed of Orbitrap analyzers and eventually the number of identified peptides. Those dependencies are not necessarily linear and can depend on many other parameters such as the sample complexity or the length of a gradient. Knowledge about these dependencies is critical to optimize the performance of the instrument. The same goes for any biochemical experiment or biological system such as the isolation efficiency of proteins or the optimal growth condition for primary cells. The conventional optimization approach is to change one parameter at a time and monitor the response of the system to titrate the optimal condition such as the number of proteins or growth rate. However, this approach is tedious, and fails to identify higher order dependencies without sampling an exponentially growing condition space.

It is apparent that sampling all potential combinations of parameters and matching these to the system response would provide a complete but unnecessarily redundant description of the system. It would be sufficient to measure some but not all points to reliably model how the system responses to individual parameter settings. Mathematical designed and evaluated experiments offer efficient sampling of the parameter space and statistically meaningful interpretation of the results. So called Design of Experiment (DoE) approaches have been increasingly popular to optimize complex systems [224]. In DoE, different mathematical models are used to approximate the real multidimensional parameter (or 'factor') space based on the evaluation of responses to defined factor settings. Different models are employed, which usually have an inverse correlation between the order of dependency that can be modeled (linear, quadratic, synergistic effects etc.) and the number of required experiments [224]. The advantages of DoE is the substantially reduced experimental effort, while allowing to assess significance of the influencing factors and to quantify higher order dependencies like synergistic effects. Figure 13 shows an example of a response surface spanned by two dimensions (two factors) and the response. In that example, non-linearity of the response would render "one parameter at a time" approaches very inefficient and error prone.
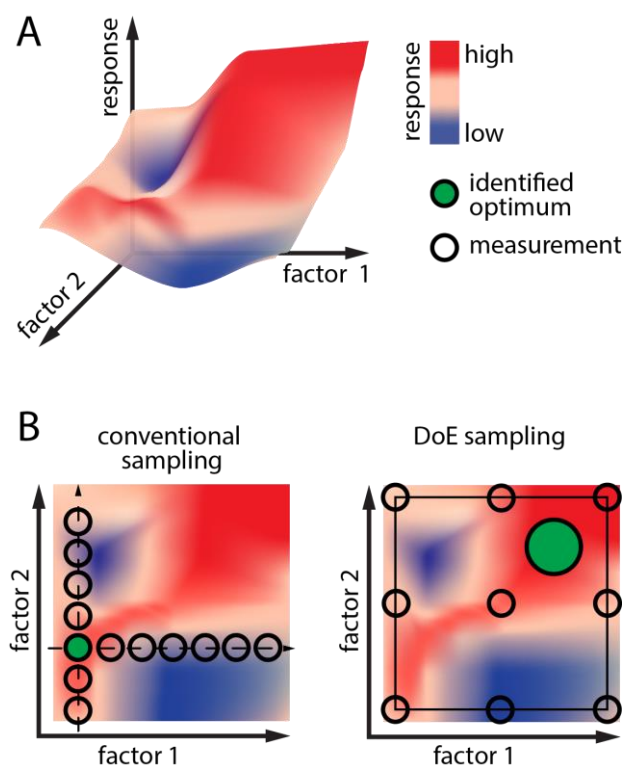

Figure 11 Comparison of conventional experimental approaches and Design of Experiment (DoE): A) A theoretical response with nonlinear dependencies on two factors. The goal is to find the parameter setting yielding the most robust and highest response. B) 2D projection of the response surface. The conventional approach — 'changing one parameter at time' — is laborious and might not find the optimal factor settings. DoE sampling, here a central composite faced design, provides better insight into the global response landscape and is more likely to provide information on the global optimum. A second DoE in the area of the determined optimum might be advisable to provide even higher resolution around the optimal settings.

A DoE usually employs three steps: (1) Identifying potentially relevant factors that might influence the system and setting realistic factor ranges. (2) Performing a screen providing information which parameters have the strongest effects. (3) Probing a subset of the relevant parameters combinations and modelling individual effects and dependencies. The resulting empirical function is usually a polynomial that allows predicting the behavior of the system. The elegance of DoE is that the responses of almost any technical and biological system can be assessed. The only critical requirements are a well quantifiable readout and adjustable factors. We used DoE to model quadratic and synergistic effects of settings affecting the number of identified peptides from the newest generation of Orbitrap hybrid MS (*article 6*).

Publications

Article 1

Deep proteomic evaluation of primary and cell line motoneuron disease models delineates major differences in neuronal characteristics

This study provides a paradigm example of how to use label-free quantitative proteomics to assess and compare properties of primary cells, tissues and cell culture systems. The fatal neurodegenerative disorder amyotrophic lateral sclerosis (ALS) primarily affects neurons of the motor system. Its cellular specificity and complex genetics (see chapter 1.3) make ALS a promising test case to investigate the network of factors contributing to neurodegeneration. To understand pathological changes, knowledge of the physiological properties of affected cells is imperative. Our collaborators have recently established a protocol to isolate and differentiate primary motoneurons from mouse embryos [225]. In this study, we characterized primary motoneurons providing a resource for motoneuron disease research. In addition, we benchmarked two *in vitro* cell line model systems (NSC-34, N2a) that are widely used to investigate (moto)neuron diseases. We compared these to mouse embryonic spinal cord tissue as well as to non-neuronal cell lines (Hepa 1-6, Mef) in order to evaluate neuronal properties. With more than 10,000 identified proteins, we were able to ascertain properties of these cell-systems in depth. We observed substantial differences between motoneuron-like cell lines and primary motoneurons, especially for proteins involved in differentiation, the cytoskeleton, and receptor signaling. However, our data also indicate that both motoneuron-like cell lines retain neuronal properties in comparison to the non-neuronal cell lines. For future research, our quantitative data will allow choosing the appropriate model system depending on the biological question to be addressed. In addition, the dataset will enable researchers to better judge the reliability of previously obtained results using the motoneuronal model cell lines.

Molecular & Cellular Proteomics (2014) [226]

Hornburg D, Drepper C, Butter F, Meissner F, Sendtner M, Mann M.

Contribution: This project was a collaboration between the department of Matthias Mann (MPI Biochemistry) and the group of Michael Sendtner (University of Wurzburg). To quantify proteomes of different model systems, I conceived the experimental setup, prepared the samples, acquired and analyzed the data and performed the follow up experiments. Our collaborators provided primary motoneurons and spinal cord tissue and contributed to data interpretation.

Deep Proteomic Evaluation of Primary and Cell Line Motoneuron Disease Models Delineates Major Differences in Neuronal Characteristics*⑤

Daniel Hornburg‡, Carsten Drepper§¶, Falk Butter‡‖, Felix Meissner‡**, Michael Sendtner¶**, and Matthias Mann‡**

The fatal neurodegenerative disorders amyotrophic lateral sclerosis and spinal muscular atrophy are, respectively, the most common motoneuron disease and genetic cause of infant death. Various *in vitro* model systems have been established to investigate motoneuron disease mechanisms, in particular immortalized cell lines and primary neurons. Using quantitative mass-spectrometry-based proteomics, we compared the proteomes of primary motoneurons to motoneuron-like cell lines NSC-34 and N2a, as well as to non-neuronal control cells, at a depth of 10,000 proteins. We used this resource to evaluate the suitability of murine *in vitro* model systems for cell biological and biochemical analysis of motoneuron disease mechanisms. Individual protein and pathway analysis indicated substantial differences between motoneuron-like cell lines and primary motoneurons, especially for proteins involved in differentiation, cytoskeleton, and receptor signaling, whereas common metabolic pathways were more similar. The proteins associated with amyotrophic lateral sclerosis also showed distinct differences between cell lines and primary motoneurons, providing a molecular basis for understanding fundamental alterations between cell lines and neurons with respect to neuronal pathways with relevance for disease mechanisms. Our study provides a proteomics resource for motoneuron research and presents a paradigm of how mass-spectrometry-based proteomics can be used to evaluate disease model systems.   *Molecular & Cellular Proteomics 13: 10.1074/mcp.M113.037291, 3410–3420, 2014.*

Motoneurons are extremely extended neurons that mediate the control of all muscle types by the central nervous system. Therefore, diseases involving progressive motoneuron degeneration such as amyotrophic lateral sclerosis (ALS)[1] (OMIM: 105400) or spinal muscle atrophy (OMIM: 253300) are particularly devastating and generally fatal disorders. Today, ALS is believed to form a phenotypic continuum with the disease entity frontotemporal lobe degeneration (OMIM: 600274) (1, 2). About 10% of ALS cases are known to be inherited, but the vast majority are considered sporadic. The number of inherited cases might be underestimated because of incomplete family histories, non-paternity, early death of family members, or incomplete penetrance (3).

Mutations in several genes have been reported for the familial form, including in *Sod1* (4), *Als2* (5), *Setx* (6), *Vapb* (7), *Tardbp* (8, 9), *Fus/Tls* (10, 11), *Vcp* (12), *Pfn1* (13), and several others (reviewed in Ref. 14). The most frequent genetic cause of inherited ALS was recently shown to be a hexanucleotide repeat expansion in an intron of a gene of unknown function called *C9orf72* (15–17). Based on the spectrum of known mutations, several disease mechanisms for ALS have been proposed, including dysfunction of protein folding, axonal transport, RNA splicing, and metabolism (reviewed in Refs. 14, 18, and 19). Despite intensive research, it is still unclear whether a main common molecular pathway or mechanism underlies motoneuron degeneration in ALS and frontotemporal lobe degeneration. Spinal muscle atrophy is caused by homozygous mutations or deletions in the survival of motor neuron gene (*Smn1*) that presumably impair the RNA metabolism through diminished functionality of the *Smn1* gene product (20). Over recent decades several model systems have been established to investigate ALS (21). These include transgenic animal models such as mouse (22), drosophila (23), and zebrafish (24). In cell-based studies, primary motoneurons cultured from rodent embryos (25) or motoneuron-

like cell lines are employed. Primary cells are considered to more closely mimic the *in vivo* situation, but they are more challenging to establish and maintain. In contrast, the degree of functional relevance of cell lines can be difficult to establish, but they can be propagated without limitation and are well suited for high-throughput analysis. In particular, the spinal cord neuron–neuroblastoma hybrid cell line NSC-34 (26) and the mouse neuroblastoma cell line N2a (27) are widely used not only to assess motoneuron function, but also to study disease mechanisms in motoneurons (28, 29).

As proteins are the functional actors in cells, proteomics should be able to make important contributions to the characterization and evaluation of cellular models. In particular, by identifying and quantifying the expressed proteins and bioinformatically interpreting the results, one can obtain enough information to infer functional differences. Our laboratory has previously shown proof of concept of such an approach by comparing the expression levels of about 4,000 proteins between primary hepatocytes and a hepatoma cell line (30). Very recently, mass-spectrometry-based proteomics has achieved sufficient depth and accuracy to quantify almost the entire proteome of mammalian cell lines (31–33). Furthermore, new instrumentation and algorithms now make it possible to perform label-free quantification between multiple cellular systems and with an accuracy previously associated only with stable isotope labeling techniques (34, 35).

To evaluate the suitability of motoneuron-like cell lines as cellular model systems for research on ALS and related disorders, we characterized the proteomes of two widely used cell lines, NSC-34 and N2a, and compared them with the proteomes of mouse primary motoneurons and non-neuronal control cell lines. To generate primary motoneurons, we employed a recently described culturing system that makes it possible to isolate highly enriched motoneuron populations in less than 8 h (25). We identified more than 10,000 proteins and investigated differences in quantitative levels of individual neuron-associated proteins and pathways related to motoneuron function and disease mechanisms.

EXPERIMENTAL PROCEDURES

Cell Culture—N2a (CCL-131, American Type Culture Collection), NSC-34 (CED-CLU140, Biozol, Eching, Germany), mouse embryonic fibroblasts (MEFs) (American Type Culture Collection), and mouse hepatoma (liver cancer; Hepa1–6) (CRL-1830, American Type Culture Collection) cell lines were grown in high-glucose Dulbecco's modified Eagle's medium with L-glutamine (PAA/GE, Freiburg, Germany), 10% FCS (PAA), 1% penicillin/streptomycin (PAA). NSC-34 cells were directly ordered from the supplier and kept in culture for up to six passages. All other non-neuronal cell lines are maintained in cryostocks in our laboratory. Cells were harvested with cold PBS. Washed pellets were flash-frozen in liquid nitrogen for storage.

Primary Motoneurons—CD1 wild-type mice were directly purchased from Charles River (Sulzfeld, Germany). Primary spinal cord motoneurons from mice were cultured essentially as described previously (25). Briefly, spinal cords of E12.5 mouse embryos were dissected and collected in 1.5-ml tubes containing Hank's Balanced Salt Solution, Life Technologies GmbH, Darmstadt Germany After

trypsin treatment and trituration, the suspension was enriched for p75^{NTR}-positive cells via an antibody-panning procedure (25). After washing and depolarization to remove the cells from the panning dish, the cells were collected via centrifugation at 400g for 5 min and resuspended in full medium (Neurobasal (Invitrogen), 1× GlutaMax (Invitrogen), supplemented with 1× B27 (Invitrogen), 2% horse serum (Linaris, Dossenheim, Germany)). After counting with a hemocytometer, 7×10^5 cells were plated on poly-ornithine- (Sigma) and laminin-coated (Invitrogen) dishes (six-well format, Delta surface, Nunc/Thermo Fisher Scientific Inc., Waltham, MA) and fed with full medium supplemented with 5 ng/ml BDNF. 40% of the medium was replaced after 24 h and every second day after that. In total, cells from ~140 CD1 wild-type embryos were used in three experiments. After 7 days in culture, cells were washed three times with PBS and harvested in 0.5 ml of lysis buffer (4% SDS, 0.1 M DTT, 10 mM Hepes, pH 8.0) per well.

Sample Preparation for MS Analysis—Cell lysis was performed at room temperature in lysis buffer (4% SDS, 10 mM Hepes, pH 8.0) with 15 min of sonication (level 5, Bioruptor, Diagenode, Seraing (Ougrée) - Belgium). Proteins in the lysate were reduced with 10 mM DTT for 30 min and then subjected to 45 min of alkylation with 55 mM iodoacetamide. The protein concentration was determined with fluorescence at 350 nm and compared with a bovine serum albumin standard.

To remove the detergent, we performed acetone precipitation. Acetone (−20 °C) was added to 100 μg of proteins to a final concentration of 80% v/v, and proteins were precipitated for at least 2 h at −20 °C. The supernatant was removed after 15 min of centrifugation (4 °C, 16,000g) followed by washing with 80% acetone (−20 °C). The protein pellet was dissolved in 50 μl of 6 M urea/2 M thiourea, 10 mM Hepes, pH 8.0. After the addition of 1 μg of LysC, digestion was carried out for 3 h at room temperature. We diluted with 4 volumes of 50 mM ammonium bicarbonate and further digested the mixture with 1 μg of trypsin overnight at room temperature. The resulting peptide mixtures were either analyzed directly via a single-shot strategy (36) or separated into six strong anion exchange fractionations as described elsewhere (37). We desalted peptides on C18 StageTips (38).

LC-MS/MS—We separated peptides on a Thermo Scientific EASY-nLC 1000 HPLC system (Thermo Fisher Scientific, Odense, Denmark). Columns (75-μm inner diameter, 50-cm length) were in-house packed with 1.9-μm C18 particles (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany). Peptides were loaded in buffer A (0.5% formic acid) and separated with a gradient from 7% to 60% buffer B (80% acetonitrile, 0.5% formic acid) within 3.5 h at 200 nl/min. The column temperature was set to 40 °C. A quadrupole Orbitrap mass spectrometer (34) (Q Exactive, Thermo Fisher Scientific) was directly coupled to the liquid chromatograph via a nano-electrospray source. The Q Exactive was operated in a data-dependent mode. The survey scan range was set to 300 to 1,650 m/z, with a resolution of 70,000 at m/z 200. Up to the 10 most abundant isotope patterns with a charge of ≥2 were subjected to Higher-energy collisional dissociation (39) with a normalized collision energy of 25, an isolation window of 2 Th, and a resolution of 17,500 at m/z 200. To limit repeated sequencing, dynamic exclusion of sequenced peptides was set to 30 s. Thresholds for ion injection time and ion target value were set to 20 ms and 3×10^6 for the survey scans and to 60 ms and 10^6 for the MS/MS scans. Data were acquired using Xcalibur software (Thermo Scientific).

Data Analysis and Statistics—To process MS raw files, we employed MaxQuant software (v. 1.3.10.18) (40). We used Andromeda (41), which is integrated into MaxQuant, to search MS/MS spectra against the UniProtKB FASTA database (59,345 forward entries; version from June 2012). Enzyme specificity was set to trypsin allowing cleavage N-terminal to proline and up to two miscleavages. Peptides had to have a minimum length of seven amino acids to be considered for identification. Carbamidomethylation was set as a fixed modifica-

tion, and acetylation (N terminus) and methionine oxidation were set as variable modifications. A false discovery rate (FDR) cutoff of 1% was applied at the peptide and protein levels. Initial precursor mass deviation of up to 4.5 ppm and fragment mass deviation up to 20 ppm were allowed. Precursor ion mass accuracy was improved by time-dependent recalibration algorithms in MaxQuant. The cutoff score (delta score) for accepting individual MS/MS spectra was 17.

Nonlinear retention time alignment of all measured samples was performed in MaxQuant. "Match between runs," which allows the transfer of peptide identifications in the absence of sequencing, was enabled with a maximum retention time window of 1 min. As a library for "match between runs" in MaxQuant, we employed additional sets of NSC-34, technical duplicates of E12.5 spinal cord, and technical duplicates of strong anion exchange fractionated peptides from motoneurons in addition to the biological samples analyzed here (biological triplicates of NSC-34, N2a, MEF, Hepa 1–6, motoneurons, and E12.5 spinal cord) (supplemental Table S6). Added individual peptide intensities are a rough proxy of each protein's abundance. In order to ensure improved quantification, we applied further normalization algorithms built into MaxQuant before making quantitative comparisons of the cellular systems. Furthermore, we stringently filtered our data by requiring a minimum peptide ratio count of 2 in MaxLFQ (42). Protein identification required at least one razor peptide (40). Proteins that could not be discriminated on the basis of unique peptides were grouped into protein groups. Furthermore, proteins were filtered for common contaminants ($n = 247$ (41)) to which laminin sequences were added to avoid bias resulting from the laminin-coating of plates in the case of the primary motoneuron culture. Proteins identified only by site modification were excluded from further analysis. Furthermore, we required a minimum of two valid MaxLFQ quantifications in at least one group of triplicates. A flowchart describing the individual steps of the data processing and filtering is presented in supplemental Table S5.

To quantify fold changes of proteins across samples, we used MaxLFQ. To visualize these fold changes in the context of individual protein abundances in the proteome, we projected them onto the summed peptide intensities normalized by the number of theoretically observable peptides (43).

Specifically, to compare relative protein abundances between and within samples, the following transformations to the MaxQuant output data were performed. First log 2 MaxLFQ differences to the mean log 2 MaxLFQ intensity of each protein group for each sample were calculated, resulting in intensity-independent log 2 MaxLFQ differences. Then protein length normalized log 2 protein intensities (termed iBAQ values in MaxQuant) were added to the MaxLFQ differences. The first transformation provides highly accurate information about the protein differences between samples, and the second allows one to compare the abundances of different protein groups.

For statistical and bioinformatic analysis, as well as for visualization, we used the open PERSEUS environment, which is part of MaxQuant. For several calculations and plots we also used the R framework (44). Pearson correlation was performed on filtered MaxLFQ intensities omitting missing values (NaN). To identify the most discriminating proteins in groupwise comparisons, first we performed imputation of missing values with a normal distribution (width = 0.3; shift = 1.8) as described elsewhere (45). Then, we applied the built-in analysis-of-variance functions in PERSEUS using an FDR of 1% and S_0 of 4 (the S_0 parameter sets a threshold for minimum fold change (46)). For pairwise comparison of proteomes and determination of significant differences in protein abundances, t test statistics were applied with a permutation-based FDR of 2% and S_0 of 4 (46). The resulting significant outliers for each of the sample pairs were analyzed for annotation enrichments.

We used outlier populations derived by means of t test statistics in order to determine Gene Ontology (GO) (47), KEGG (48), and UniProt Keyword (49) enrichment of categorical annotations with an intersection size of at least 8. In Fig. 3 this was done by testing categorical annotations in individual outlier populations for enrichments compared with the complete dataset via Fisher's exact test with a Benjamini–Hochberg corrected FDR of 2% (Fig. 3C). In Fig. 4, no FDR cutoff was used, but the FDR values are reported (supplemental Table S3). In cases of identical annotation in both outlier populations, only the enrichment with highest p value was depicted.

For comparison of the motoneuronal proteome to published transcriptome data (50), we used normalized and filtered log 2 Affymetrix intensities provided by the authors for transcripts and protein intensities from MaxQuant, which were normalized by protein sequence length. Transcriptome data were matched in PERSEUS to proteome data on the basis of gene names, and the median of replicates was used for comparison.

RESULTS

Deep Proteome Analysis of Motoneuronal Model Systems—We set out to characterize the proteomes of different cellular model systems employed in the study of motoneuron disease mechanisms. Primary motoneurons were isolated from 12.5-day-old mouse embryos, enriched and differentiated using a recently developed protocol (25), and compared with two commonly used motoneuron-like cell lines (NSC-34 and N2a) and spinal cord from 12.5-day-old mouse embryos. As non-neuronal control cells, we employed mouse hepatoma cells (Hepa 1–6) and MEF cells. To ensure optimal recovery of membrane proteins such as receptors, which are crucial for neuronal function, we extracted proteins from cells via lysis in SDS followed by acetone precipitation. After enzymatic digestion, tryptic peptides were separated by means of reverse phase liquid chromatography in 4 h gradients and analyzed on a quadrupole Orbitrap mass spectrometer in single shots (Fig. 1A). Label-free quantification between biological replicates was highly reproducible, as evidenced by a Pearson correlation of more than 0.93 (Fig. 1B). Overall we quantified more than 10,000 proteins and 7,000 to 8,100 proteins in the individual cellular systems (Fig. 1C, supplemental Tables S1 and S2).

About 75% of all quantified proteins were present in all analyzed cellular systems (primary cells, motoneuronal cell lines, and non-neuronal controls). We observed the most pairwise overlap of detected proteins between primary neuronal cells and motoneuronal cell lines (Fig. 1D). However, on a quantitative level, motoneuronal cell lines correlated more tightly to the non-neuronal controls than to the primary cells. This unexpected observation highlights the value of quantitative data (Fig. 1B). Measured protein abundance values (summed, normalized intensities of the peptides identifying each protein) covered over 6 orders of magnitude (Fig. 1E). In all cell systems, some of the most abundant proteins were histones and ribosomal proteins, in accordance with previous studies (43, 51, 52). Proteins of housekeeping metabolic processes such as spliceosomes or oxidative phosphorylation were also generally highly abundant. In contrast, neuronal
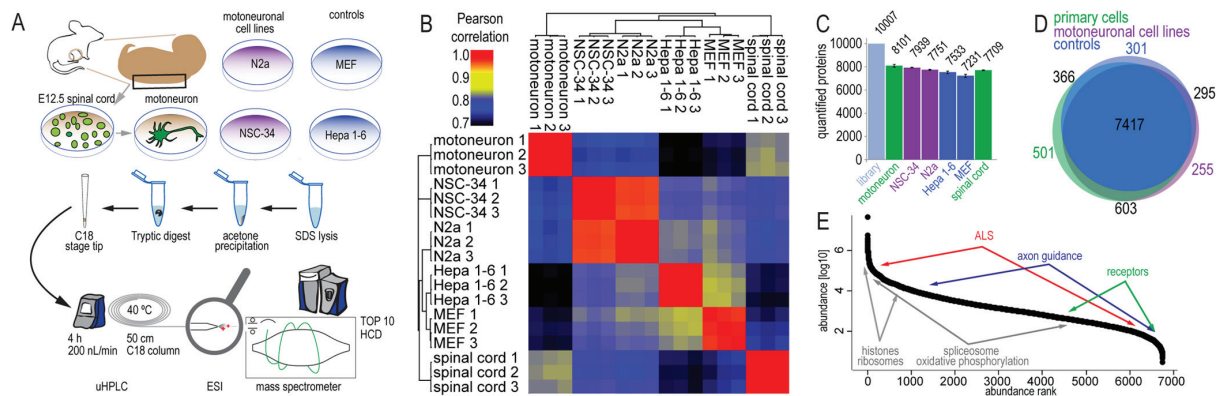
FIG. 1. **Deep proteome analysis of motoneuronal model systems.** *A*, spinal cord from E12.5 mouse embryo was dissected, and cells were isolated. P75[NTR]-positive cells were enriched in an antibody-panning step and grown in culture for 7 days. Motoneuron-like cell lines (N2a, NSC-34) and control cell lines (MEF, Hepa 1–6) were grown under standard culture conditions. All investigated cellular systems were analyzed via liquid chromatography coupled to high-resolution mass spectrometry following a single-shot strategy. *B*, Pearson correlation for MS-determined protein abundances (MaxLFQ intensities) from single-shot MS measurements. All replicate correlation values were greater than 0.93, and Pearson correlation between different cellular systems was at least 0.7. *C*, total number of quantified proteins (summed peptide intensities) in all measurements, including additional replicates of cell lines and spinal cord ("library") and in individual systems. Between 7,000 and 8,100 proteins were quantified in the individual cellular systems. *D*, Venn diagram of the distribution of quantified proteins in the investigated cellular systems, grouped by primary cells, motoneuron-like cell lines, and non-neuronal cell lines. *E*, ranked MaxLFQ normalized protein abundances for motoneurons. Indicated are approximated abundance ranges for proteins involved in neuron-specific and general metabolic processes and in ALS pathology.
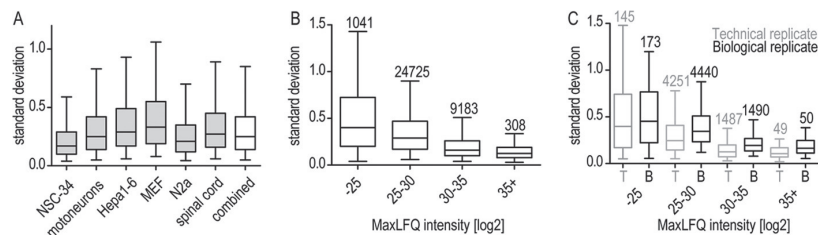

FIG. 2. **Standard deviation of MaxLFQ quantifications in protein groups within replicates.** *A*, standard deviation of MaxLFQ quantifications in protein groups for triplicates. *B*, intensity dependence of MaxLFQ standard deviations for all triplicates. The number of data points contributing to each bin is depicted above the bins. *C*, comparison of standard deviations from technical duplicates and biological triplicates of spinal cord throughout the whole abundance range. Differences between technical and biological replicates are significant ($p < 0.0001$) for protein intensities of bins 25–30, 30–35, and 35+. The number of protein groups for which standard deviations were calculated is depicted above the bins. Data are plotted as median with 5%–95% percentile.

proteins involved in receptor signaling or axonal guidance were relatively low abundant. Because ALS-associated proteins are involved in a variety of cellular processes, they span almost the entire dynamic range (supplemental Table S2).

For selected gene products—Dctn1, Vcp, Fus, Syp, and Sox2—we performed Western blot analysis on motoneuron primary culture, NSC-34, and Hepa 1–6 cell lines. Although Western blot staining was less easy to interpret than the MS signals, trends in protein abundance were the same (supplemental Figs. S1*A*–S1*C*). We also compared our motoneuronal dataset to a published transcriptome of murine motoneurons that were differentiated from stem cells with retinoic acid and sonic hedgehog for 5 days (50) (supplemental Fig. S1*E*, supplemental Table S4). We observed a moderate and highly significant correlation (Pearson correlation, 0.49; Pearson *p* value, 10e-15) between transcriptome and proteome; how-

ever, many proteins differed strongly between the two datasets (supplemental Table S4).

Global Comparison of the Proteomes of Primary Motoneurons, Cell Lines, and Mouse Spinal Cord—After stringent filtering as described in "Experimental Procedures" and supplemental Table S5, we quantitatively compared 7,950 proteins between the model systems on the level of individual protein expression as well as annotation levels (supplemental Table S2). The mean standard deviation within biological replicates was less than 2-fold for 95% of the quantified proteins, demonstrating the quantitative robustness of our measurements (Fig. 2*A*). Even for the least abundant proteins, the mean standard deviation was less than 2-fold for more than 99.5% of these proteins (Fig. 2*B*). As an example, for spinal cord tissue, the variation between one biological sample measured twice (technical replicate) was moderately but sig-

nificantly smaller than the variation between biological replicates (Fig. 2C).

We applied principal component analysis (PCA) to evaluate similarities and dissimilarities on a global scale. Biological replicates clustered tightly in the PCA space, indicating high quantitative reproducibility of the experimental setup and the label-free quantification (Fig. 3A). Motoneurons, spinal cord, Hepa 1–6, MEFs, and motoneuronal cell lines were clearly separated in the first two components of PCA. NSC-34 and N2a together formed a close pair of clusters that was approximately equidistant from the primary cells and the control cell lines. Although derived from completely different tissues, the latter are co-located and widely separated from the other systems. As the difference between them is not profound, we focused on Hepa 1–6 as a non-neuronal cell line control for further analysis. Interestingly, the motoneuron primary culture and the E12.5 spinal cord tissue, although clearly separated, clustered much closer to each other than to the neuronal and non-neuronal cell lines.

To assess individual, discriminating proteins in motoneurons, spinal cord, N2a, NSC-34, and Hepa 1–6 cells, we plotted the "loadings" of the PCA (loadings are the proteins driving, or responsible for, the separation of the proteomes). In Fig. 3B, those loadings that also passed a stringent analysis of variance and fold change ("Experimental Procedures") are colored. Within this selection we can distinguish groups of proteins on the basis of neuronal properties. Proteins that function in neurite growth, such as Gap43 and Syn2, or in vesicle release, such as the neuro-secretory protein Vgf, were shared outliers for both neuron-like cell lines and motoneurons/spinal cord (orange population in Fig. 3B). In contrast, another group of proteins—such as medium and light chains of neurofilaments (Nefm and Nefl)—proved to be distinctive for motoneuronal systems relative to all other systems (green population in Fig. 3B).

We employed PERSEUS to bioinformatically determine major physiological differences among the cellular model systems via enriched annotations (GO, UniProt Keyword, and KEGG) (Fig. 3C, supplemental Fig. S3, supplemental Table S3). E12.5 spinal cord consists of a mixed population of undifferentiated cells. Consistent with this, annotation enrich-

Fig. 3. **Global proteomic comparison of motoneuron-like cell lines, embryonic spinal cord, unrelated cell lines, and primary motoneurons.** A, PCA of protein expression values of the different systems. Component 1 and component 2 account for 54.8% of the data variation. B, significant outliers from analysis of variance (FDR cutoff = 1%, S_0 fold-change cutoff = 4) are depicted in red in the PCA loading distribution (gray dots). Proteins driving the separation are colored according to the PCA plot in A. C, analyses of enriched annotation (Fisher's exact test) for t test significant (FDR cutoff = 2%, S_0 fold-change cutoff = 4) outlier populations. The six most enriched (log 2) annotations from UniProt Keywords (Key), GO biological process (GOBP), GO cellular compartment (GOCC), and KEGG are depicted.

ments for primary motoneurons demonstrated a stronger neuronal character relative to spinal cord (supplemental Fig. S3). NSC-34 and N2a cells differed even more from primary motoneurons with respect to neuron-specific proteins. For instance, neuron-associated annotations such as GO biological process "neurotransmitter transport" were enriched more than four times in motoneurons relative to NSC-34 and N2a (Fig. 3C). Functions that were enriched for these neuronal cell lines were mainly related to DNA metabolism, presumably as a consequence of the fact that they are proliferating whereas the primary motoneurons are terminally differentiated.

Although they are derived from neuroblastoma, the overall neuronal characteristics of NSC-34 and N2a are not profound. Nevertheless, comparison of NSC-34 and N2a cells to non-neuronal cell lines revealed some patterns that are characteristic for neurons—for instance, enrichment of proteins involved in neuron projection and the neuronal cytoskeleton. These observations based on proteomic quantification highlight fundamental differences between neuronal cell lines and primary motoneurons with respect to their protein repertoire for neuronal differentiation and function.

To characterize differences in protein patterns among primary motoneurons, motoneuronal cell lines, and non-neuronal cell lines in more detail, we computed enrichment values and *p* values for annotation categories for binary comparisons of the cellular systems (Fig. 4, supplemental Table S3). Comparison of primary motoneurons to NSC-34 and N2a cells revealed striking differences for neuronal categories, with a large overrepresentation in the motoneurons.

Evaluation of Axon Growth, Guidance, and Polarity—Impaired establishment and maintenance of neuronal morphology is a central pathological aspect in neurodegenerative disorders (53). We had observed the corresponding annotations such as neuron projection (GO cellular component) and axon guidance (KEGG) to be de-enriched in the motoneuronal cell lines, and we next wished to extract the expression level of key factors in these processes (Fig. 5A). Illustrating the use of the dataset as a resource, we observed that proteins important for axonal outgrowth like Slits/Robos, Semaphorins, Netrins, or Ephrins or proteins involved in morphological maintenance such as Mapt were either missing or underrepresented in NSC-34 and N2a.

Evaluation of Excitability and Synaptic Function—Neurotransmitter-receptor signaling and excitability are main features of neuronal cell types that differ according to cellular functionality. We assessed the abundance of proteins involved in receptor signaling (KEGG "neuroactive receptor-ligand interaction," voltage-gated channel (UniProt Keyword), and ligand-gated ion channel (UniProt Keyword); Fig. 5). Strikingly, many receptors were underrepresented or even not detected in the NSC-34 and N2a cells. In particular, glutamate receptors were present almost exclusively in isolated motoneurons, and to a lesser extent in spinal cord, but not in the neuronal cell lines. The serotonin receptor functions in fast
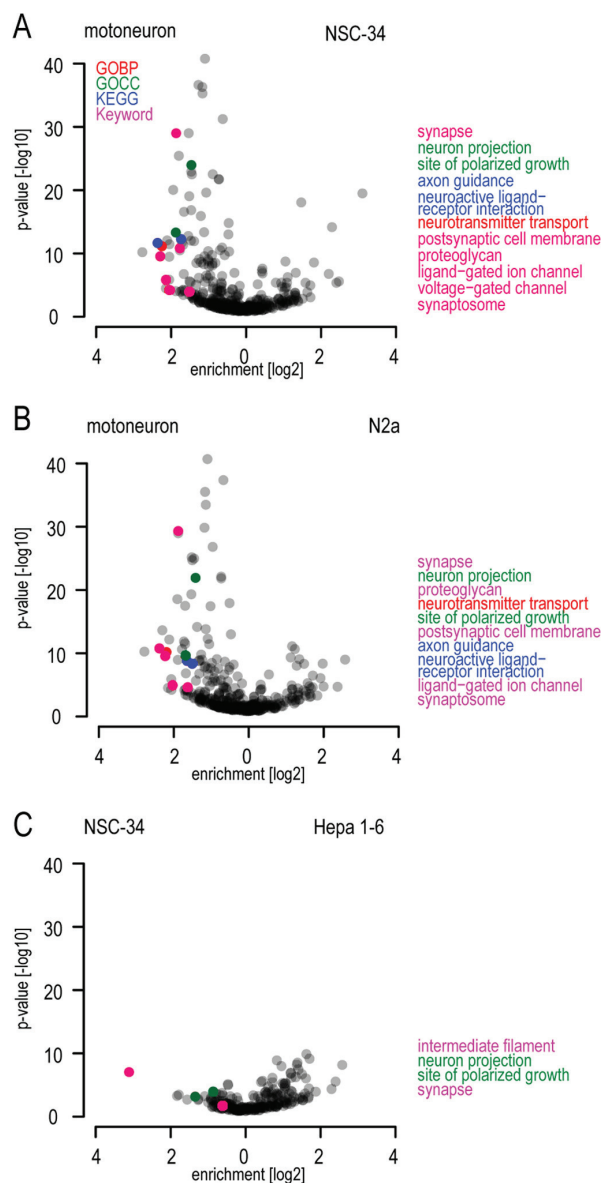

FIG. 4. Evaluation of neuronal pathway enrichments and neuron-specific protein abundances. *A–C*, pairwise comparison of enriched annotations in different cellular model systems. Annotations are depicted by their enrichment factor and statistical significance, similar to volcano plots for protein populations.

depolarization of neurons and is not expected to be present in motoneurons. Accordingly, we did not detect it in motoneurons. It was, however, identified in NSC-34 and N2a cells, suggesting that they do not faithfully represent motoneuronal characteristics in this respect. Overall, slightly more receptor proteins were present in N2a than in NSC-34 cells. Despite the striking differences of both neuronal cell lines relative to primary motoneurons, their protein abundance patterns

A — axon guidance and neuron projection

	motoneuron	NSC-34	N2a	Hepa 1-6	spinal cord
Cxcr4	5.2				5.9
Efnb3	8.2				7.4
Epha2	7.7			14	8.2
Epha4	6.6				6.5
Fyn	12.4	9.2	7.1	7.5	9.8
Mapt	16.9	11.9	13.1	8.5	14.9
Ncam1	17.9	14	14	8.2	16.4
Ngfr	7.2	10.3	10.2		11
Ntn1	8.8				7.3
Pak1	13.6	11.1	10.3	10.8	13.6
Pak2	14.1	13.2	13.1	14	14.6
Psen1	5.6				6.2
Psen2	8.4	7.2			
Rac1	16	14.4	15.6	15.4	15.3
Rac3	11.5		7		10.4
Robo1	7.8		7.5	7.4	8.1
Robo2	7.7			5.8	6.7
Robo3			5.4		6.1
Sema3a	5.1				
Sema3c	8.4	5			6.5
Sema4b	6.8	6.6		4.5	6.2
Sema4c	8	9.4	9.4	9	8.5
Sema4d		5			
Sema4g	3.7				
Sema5b	3.8				4.6
Sema7a	4.5				
Shank2	5.9		5.2		5.7
Slit1	9.8	7.2	6	6.6	8.3
Slit2	10.3	8.3	5.9	7.3	9.7
Stmn2	15.8		14.1	12.8	15.4
Stmn4	6.8				

B — synapse/ receptors

	motoneuron	NSC-34	N2a	Hepa 1-6	spinal cord
Ache	5.6				5.2
Amph	14.9	5.1			10.8
Cacna1b	6.5				5.8
Cacnb3	8.3	6.5	6.4	6.3	8.2
Calcrl				7.7	
Cdk5	12.8	11.1	10.9	11.1	10.7
Chrna4	5.8				7.9
Erc2	8.3			7	6.7
Gabbr1	11.9		6.8		10.5
Gabbr2	4.8			4.9	
Gabra2	6.5				
Gabra3	7.4				6.6
Gabrb3	9.2				6.1
Gap43	16.2	13.3	12.6		14.5
Glra2	4.6				
Gria1	8.7				
Gria2	12		7	6.7	7.7
Gria3	5.2				
Gria4	5.8				
Grid1	4.4				
Grik2	6.2				4.7
Grik3	5.3				5.2
Grin1	7.8	4.9	4.8		
Grin2b	4.7		4.5	4.6	
Grm3	6.9				
Grm5	8.9			5	6.2
Grm7	5.1				
Hcn4	3.4				
Htr3a		5.8	5.9		
Itpr1	6.2	4.8	8.3	8.6	5.7
Itpr3	6.8	5.8	6	13.2	5.1
Kctd12	14.8	8	8.2	13.8	11.4
Kctd8	7.7				6.5
Nlgn2	7.7				7.1
Nlgn3	9.2				8.2
Nlgn4l	3.8				
P2rx7		7		6.1	
Pclo	10.4	3.5	3.6		4.3
Rims1	7.8				4.4
Rims2	6.4	6.2	5.5		6.6
Rims3	6.8		6.7		
Rims4	4.1				
Ryr2	1.6				
Scn2a1	6.6	4.6			
Scn3a	2.4				
Sept11	14.8	14.6	14.7	13.6	15.3
Snap25	16.4	12.9	12.6		9.3
Stx1a	12.7	11.7	10.8		9.9
Sv2a	15.4	7.7	9.3	7.6	11.8
Syt2	8.2				5.6
Unc13a	8.7		5	3.9	4.2

abundance [log2]: 0 — 5 — 10 — 15 — 20 — 25

FIG. 5. **Evaluation of neuron-specific protein abundances.** *A, B,* abundances of selected proteins that are members of the neuronal annotation classes in Fig. 3. The normalized MS-derived abundance values allow comparisons within and between cellular systems.

clearly indicated a neuronal inheritance compared with Hepa 1–6.

Evaluation of ALS-associated Proteins—As mentioned above, ALS-associated proteins span a wide range of cellular functions and are not generally neuron specific. The presence and abundance of these ALS-associated proteins are important aspects for studying disease mechanisms and therefore for evaluating the suitability of disease models. We found that most ALS-associated proteins were indeed present in the proteomes of NSC-34 and N2a cell lines (Fig. 6). Compared with neuronal proteins, the differences between the cellular systems were smaller but still distinct. For example, whereas we observed major expression-level differences for ALS-associated neuron-specific proteins such as the Eph4 receptor, the neurofilament protein Nefl, and the microtubule-associated Mapt, levels were similar for proteins that are not neuron specific, such as Sod1.

DISCUSSION

Motoneuron diseases such as ALS or spinal muscle atrophy are incurable, and no effective symptomatic treatments are available. Researchers make use of a variety of different cell culture model systems to unravel disease pathways and to understand pathological mechanisms (19, 21, 54). However, the proteomics properties of these systems have never been assessed on a global level. Here we provide a large-scale and quantitative analysis comparing the proteomes of primary motoneurons to the neuroblastoma cell lines NSC-34 and N2a, as well as to unrelated cell lines. The quantitative inventory of expressed proteins in these systems should serve as a valuable resource for the community. We have made these data easily accessible in the MaxQB database (55). Users can query the expression rank order of proteins of interest, sequence coverage, high-resolution MS/MS spectra of peptides, and many other features of the proteome. Moreover, our study presents a generic approach with which to evaluate model systems using high-resolution quantitative mass-spectrometry-based proteomics.

Unexpectedly, our proteomics results at both global and individual protein levels only place the neuronal cell line models halfway between the actual target tissue (*in vivo* motoneurons) and completely unrelated cell lines derived from mouse liver cancer or mouse embryos. We found proteins with high relevance for neuronal function, in particular neurotransmitter receptors and proteins involved in neuronal excitability and neurotransmitter release, to be strongly underrepresented in NSC-34 and N2a. The abundances of glutamate receptors were considerably lower as well, suggesting nonexistent or at least significantly diminished neuronal excitability. Because disease pathways associated with excitotoxicity, in particular glutamate toxicity, are frequent areas of focus in ALS research (19, 56), this leads us to question the suitability of these cell lines for such studies. The nerve growth factor receptor is highly abundant during development and when neurons are lesioned or degenerating (57–60). In accordance with this, we found nerve growth factor receptor to be highly abundant in E12.5 spinal cord, whereas it was down-regulated in primary motoneurons, which are terminally differentiated. This receptor also showed relatively high abundance in NSC-34 and N2a cell lines, further suggesting differences relative to motoneurons.

Mechanisms that destabilize axons are a major focus of research in neurodegeneration (53). In this regard it is impor-
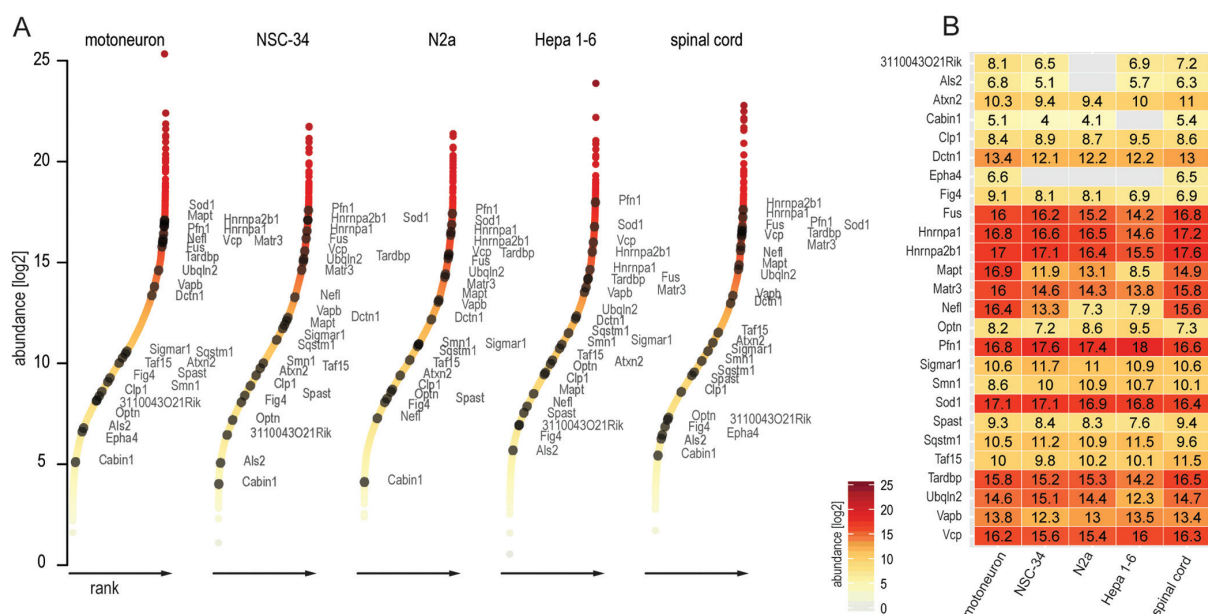
FIG. 6. **Evaluation of ALS-associated protein abundances.** *A*, ranked abundances of ALS-associated proteins in the context of the entire proteome. *B*, normalized MS-derived expression values of the same proteins as in *A*, depicted as a heat map for motoneurons, NSC-34, N2a, Hepa 1–6, and E12.5 spinal cord.

tant to take the observed differences for cytoskeleton-associated proteins—for instance, for Mapt and Nefl, which were much lower in N2a and NSC-34 cells than in primary motoneurons—into account. In addition, various members of the axon guidance family such as Slits/Robos, Semaphorins, Netrins, or Ephrins showed altered abundance patterns in NSC-34 and N2a. Together, these observations raise the question of whether N2a and NSC-34 cells are reliable models regarding motoneuronal structure, in particular axon growth and maintenance. Although both cell lines certainly exhibit aspects of neuronal morphology, at the proteomics level, neuronal features mainly become apparent in comparison to unrelated, non-neuronal cell lines. Our data thus indicate that the morphological tendency toward motoneurons does not correspond directly to the protein level. This also suggests that morphological comparisons can usefully be complemented by quantitative proteomic data.

In contrast to these motoneuron-specific pathways, universal metabolic and housekeeping pathways were less altered between cell lines and primary cells. The main difference highlighted by bioinformatic analysis of the proteomic data was an up-regulation of DNA metabolism (Fig. 3). This is because cell culture favors rapidly dividing cells, and any non-essential metabolic pathways are counter-selected. The same reasoning might also explain the down-regulation of pathways such as receptor signaling and of proteins important for the maintenance of complex cellular morphologies noted above.

We further found differences in the expression levels of some, but not all, ALS-associated proteins in the analyzed cellular systems. For many of them, such as mutant Sod1 or wild-type Tdp-43 (61), even relatively small expression levels (2-fold overexpression relative to wild-type levels, for instance (62)) are known to have severe effects on the pathogenic cascade and the onset of symptoms (63). Furthermore, motoneuron diseases are highly cell-type specific, suggesting a narrow physiological niche for manifestation of a pathological phenotype. Therefore, in addition to the fact that moderate differences in ALS protein abundances may be very important, the general motoneuronal characteristics as evidenced by quantitative proteomics need to be taken into account. Thus neuroblastoma cell lines are limited in terms of their suitability to serve as model systems for motoneurons for the study of ALS-relevant functional mechanisms, as they neither reconstitute global motoneuronal properties nor faithfully mimic motoneurons on the individual protein level. However, they still retain neuronal character in comparison to other cell lines, rendering them a valid choice for some biological investigations and high-throughput screens.

In conclusion, we here present a resource for motoneuron disease research. Because of the high protein coverage in cell lines and primary motoneurons, our data highlight important cellular pathways unique to or enriched in neurons and functionally relevant in motoneurons. Our study exemplifies how quantitative proteomics can be used to critically evaluate the suitability of model systems in general. De-

pending on the intended experiments, the best cellular model system can be chosen according to the presence or abundance of the involved receptors or neuronal pathways (such as ALS or signaling studies). In the current context, for instance, it would be interesting to compare the proteome of primary motoneurons to *in vitro* differentiated motoneurons derived from induced patient-derived pluripotent stem cells with a view to evaluate these exciting and promising new model systems.

REFERENCES

1. Lillo, P., and Hodges, J. R. (2009) Frontotemporal dementia and motor neurone disease: overlapping clinic-pathological disorders. *J. Clin. Neurosci.* **16,** 1131–1135

2. Mackenzie, I. R., Rademakers, R., and Neumann, M. (2010) TDP-43 and FUS in amyotrophic lateral sclerosis and frontotemporal dementia. *Lancet Neurol.* **9,** 995–1007

3. Al-Chalabi, A., Fang, F., Hanby, M. F., Leigh, P. N., Shaw, C. E., Ye, W., and Rijsdijk, F. (2010) An estimate of amyotrophic lateral sclerosis heritability using twin data. *J. Neurol. Neurosurg.* **81,** 1324–1326

4. Rosen, D. R. (1993) Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* **364,** 362

5. Hadano, S., Hand, C. K., Osuga, H., Yanagisawa, Y., Otomo, A., Devon, R. S., Miyamoto, N., Showguchi-Miyata, J., Okada, Y., Singaraja, R., Figlewicz, D. A., Kwiatkowski, T., Hosler, B. A., Sagie, T., Skaug, J., Nasir, J., Brown, R. H., Scherer, S. W., Rouleau, G. A., Hayden, M. R., and Ikeda, J. E. (2001) A gene encoding a putative GTPase regulator is mutated in familial amyotrophic lateral sclerosis 2. *Nat. Genet.* **29,** 166–173

6. Chen, Y. Z., Bennett, C. L., Huynh, H. M., Blair, I. P., Puls, I., Irobi, J., Dierick, I., Abel, A., Kennerson, M. L., Rabin, B. A., Nicholson, G. A., Auer-Grumbach, M., Wagner, K., De Jonghe, P., Griffin, J. W., Fischbeck, K. H., Timmerman, V., Cornblath, D. R., and Chance, P. F. (2004) DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). *Am. J. Hum. Genet.* **74,** 1128–1135

7. Nishimura, A. L., Mitne-Neto, M., Silva, H. C., Richieri-Costa, A., Middleton, S., Cascio, D., Kok, F., Oliveira, J. R., Gillingwater, T., Webb, J., Skehel, P., and Zatz, M. (2004) A mutation in the vesicle-trafficking protein VAPB causes late-onset spinal muscular atrophy and amyotrophic lateral sclerosis. *Am. J. Hum. Genet.* **75,** 822–831

8. Neumann, M., Sampathu, D. M., Kwong, L. K., Truax, A. C., Micsenyi, M. C., Chou, T. T., Bruce, J., Schuck, T., Grossman, M., Clark, C. M., McCluskey, L. F., Miller, B. L., Masliah, E., Mackenzie, I. R., Feldman, H., Feiden, W., Kretzschmar, H. A., Trojanowski, J. Q., and Lee, V. M. (2006) Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314,** 130–133

9. Sreedharan, J., Blair, I. P., Tripathi, V. B., Hu, X., Vance, C., Rogelj, B., Ackerley, S., Durnall, J. C., Williams, K. L., Buratti, E., Baralle, F., de Belleroche, J., Mitchell, J. D., Leigh, P. N., Al-Chalabi, A., Miller, C. C., Nicholson, G., and Shaw, C. E. (2008) TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science* **319,** 1668–1672

10. Vance, C., Rogelj, B., Hortobagyi, T., De Vos, K. J., Nishimura, A. L., Sreedharan, J., Hu, X., Smith, B., Ruddy, D., Wright, P., Ganesalingam, J., Williams, K. L., Tripathi, V., Al-Saraj, S., Al-Chalabi, A., Leigh, P. N., Blair, I. P., Nicholson, G., de Belleroche, J., Gallo, J. M., Miller, C. C., and Shaw, C. E. (2009) Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science* **323,** 1208–1211

11. Kwiatkowski, T. J., Jr., Bosco, D. A., Leclerc, A. L., Tamrazian, E., Vanderburg, C. R., Russ, C., Davis, A., Gilchrist, J., Kasarskis, E. J., Munsat, T., Valdmanis, P., Rouleau, G. A., Hosler, B. A., Cortelli, P., de Jong, P. J., Yoshinaga, Y., Haines, J. L., Pericak-Vance, M. A., Yan, J., Ticozzi, N., Siddique, T., McKenna-Yasek, D., Sapp, P. C., Horvitz, H. R., Landers, J. E., and Brown, R. H., Jr. (2009) Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science* **323,** 1205–1208

12. Johnson, J. O., Mandrioli, J., Benatar, M., Abramzon, Y., Van Deerlin, V. M., Trojanowski, J. Q., Gibbs, J. R., Brunetti, M., Gronka, S., Wuu, J., Ding, J. H., McCluskey, L., Martinez-Lage, M., Falcone, D., Hernandez, D. G., Arepalli, S., Chong, S., Schymick, J. C., Rothstein, J., Landi, F., Wang, Y. D., Calvo, A., Mora, G., Sabatelli, M., Monsurro, M. R., Battistini, S., Salvi, F., Spataro, R., Sola, P., Borghero, G., Galassi, G., Scholz, S. W., Taylor, J. P., Restagno, G., Chio, A., Traynor, B. J., and Consortium, I. (2010) Exome sequencing reveals VCP mutations as a cause of familial ALS. *Neuron* **68,** 857–864

13. Wu, C. H., Fallini, C., Ticozzi, N., Keagle, P. J., Sapp, P. C., Piotrowska, K., Lowe, P., Koppers, M., McKenna-Yasek, D., Baron, D. M., Kost, J. E., Gonzalez-Perez, P., Fox, A. D., Adams, J., Taroni, F., Tiloca, C., Leclerc, A. L., Chafe, S. C., Mangroo, D., Moore, M. J., Zitzewitz, J. A., Xu, Z. S., van den Berg, L. H., Glass, J. D., Siciliano, G., Cirulli, E. T., Goldstein, D. B., Salachas, F., Meininger, V., Rossoll, W., Ratti, A., Gellera, C., Bosco, D. A., Bassell, G. J., Silani, V., Drory, V. E., Brown, R. H., Jr., and Landers, J. E. (2012) Mutations in the profilin 1 gene cause familial amyotrophic lateral sclerosis. *Nature* **488,** 499–503

14. Robberecht, W., and Philips, T. (2013) The changing scene of amyotrophic lateral sclerosis. *Nat. Rev. Neurosci.* **14,** 248–264

15. Renton, A. E., Majounie, E., Waite, A., Simón-Sánchez, J., Rollinson, S., Gibbs, J. R., Schymick, J. C., Laaksovirta, H., van Swieten, J. C., Myllykangas, L., Kalimo, H., Paetau, A., Abramzon, Y., Remes, A. M., Kaganovich, A., Scholz, S. W., Duckworth, J., Ding, J., Harmer, D. W., Hernandez, D. G., Johnson, J. O., Mok, K., Ryten, M., Trabzuni, D., Guerreiro, R. J., Orrell, R. W., Neal, J., Murray, A., Pearson, J., Jansen, I. E., Sondervan, D., Seelaar, H., Blake, D., Young, K., Halliwell, N., Callister, J. B., Toulson, G., Richardson, A., Gerhard, A., Snowden, J., Mann, D., Neary, D., Nalls, M. A., Peuralinna, T., Jansson, L., Isoviita, V.-M., Kaivorinne, A.-L., Hölttä-Vuori, M., Ikonen, E., Sulkava, R., Benatar, M., Wuu, J., Chiò, A., Restagno, G., Borghero, G., Sabatelli, M., Heckerman, D., Rogaeva, E., Zinman, L., Rothstein, J. D., Sendtner, M., Drepper, C., Eichler, E. E., Alkan, C., Abdullaev, Z., Pack, S. D., Dutra, A., Pak, E., Hardy, J., Singleton, A., Williams, N. M., Heutink, P., Pickering-Brown, S., Morris, H. R., Tienari, P. J., and Traynor, B. J. (2011) A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72,** 257–268

16. DeJesus-Hernandez, M., Mackenzie, I. R., Boeve, B. F., Boxer, A. L., Baker, M., Rutherford, N. J., Nicholson, A. M., Finch, N. A., Flynn, H., Adamson, J., Kouri, N., Wojtas, A., Sengdy, P., Hsiung, G. Y., Karydas, A., Seeley, W. W., Josephs, K. A., Coppola, G., Geschwind, D. H., Wszolek, Z. K., Feldman, H., Knopman, D. S., Petersen, R. C., Miller, B. L., Dickson, D. W., Boylan, K. B., Graff-Radford, N. R., and Rademakers, R. (2011) Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72,** 245–256

17. May, S., Hornburg, D., Schludi, M. H., Arzberger, T., Rentzsch, K., Schwenk, B. M., Grasser, F. A., Mori, K., Kremmer, E., Banzhaf-Strathmann,

J., Mann, M., Meissner, F., and Edbauer, D. (2014) C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. *Acta Neuropathol.* **128,** 485–503

18. Al-Chalabi, A., Jones, A., Troakes, C., King, A., Al-Sarraj, S., and van den Berg, L. H. (2012) The genetics and neuropathology of amyotrophic lateral sclerosis. *Acta Neuropathol,* **124,** 339–352

19. Ferraiuolo, L., Kirby, J., Grierson, A. J., Sendtner, M., and Shaw, P. J. (2011) Molecular pathways of motor neuron injury in amyotrophic lateral sclerosis. *Nat. Rev. Neurol.* **7,** 616–630

20. Fallini, C., Bassell, G. J., and Rossoll, W. (2012) Spinal muscular atrophy: the role of SMN in axonal mRNA regulation. *Brain Res.* **1462,** 81–92

21. O'Connor, D. M., and Boulis, N. M. (2012) Cellular and molecular approaches to motor neuron therapy in amyotrophic lateral sclerosis and spinal muscular atrophy. *Neurosci. Lett.* **527,** 78–84

22. Gurney, M. E., Pu, H. F., Chiu, A. Y., Dalcanto, M. C., Polchow, C. Y., Alexander, D. D., Caliendo, J., Hentati, A., Kwon, Y. W., Deng, H. X., Chen, W. J., Zhai, P., Sufit, R. L., and Siddique, T. (1994) Motor-neuron degeneration in mice that express a human Cu,Zn superoxide-dismutase mutation. *Science* **264,** 1772–1775

23. Tsuda, H., Han, S. M., Yang, Y. F., Tong, C., Lin, Y. Q., Mohan, K., Haueter, C., Zoghbi, A., Harati, Y., Kwan, J., Miller, M. A., and Bellen, H. J. (2008) The amyotrophic lateral sclerosis 8 protein VAPB is cleaved, secreted, and acts as a ligand for Eph receptors. *Cell* **133,** 963–977

24. Laird, A. S., Van Hoecke, A., De Muynck, L., Timmers, M., Van den Bosch, L., Van Damme, P., and Robberecht, W. (2010) Progranulin is neurotrophic in vivo and protects against a mutant TDP-43 induced axonopathy. *PLoS One* **5,** e13368

25. Wiese, S., Herrmann, T., Drepper, C., Jablonka, S., Funk, N., Klausmeyer, A., Rogers, M. L., Rush, R., and Sendtner, M. (2010) Isolation and enrichment of embryonic mouse motoneurons from the lumbar spinal cord of individual mouse embryos. *Nat. Protoc.* **5,** 31–38

26. Cashman, N. R., Durham, H. D., Blusztajn, J. K., Oda, K., Tabira, T., Shaw, I. T., Dahrouge, S., and Antel, J. P. (1992) Neuroblastoma x spinal cord (NSC) hybrid cell lines resemble developing motor neurons. *Dev. Dyn.* **194,** 209–221

27. Oh, Y. K., Shin, K. S., Yuan, J., and Kang, S. J. (2008) Superoxide dismutase 1 mutants related to amyotrophic lateral sclerosis induce endoplasmic stress in neuro2a cells. *J. Neurochem.* **104,** 993–1005

28. Matusica, D., Fenech, M. P., Rogers, M. L., and Rush, R. A. (2008) Characterization and use of the NSC-34 cell line for study of neurotrophin receptor trafficking. *J. Neurosci. Res.* **86,** 553–565

29. Eggett, C. J., Crosier, S., Manning, P., Cookson, M. R., Menzies, F. M., McNeil, C. J., and Shaw, P. J. (2000) Development and characterisation of a glutamate-sensitive motor neurone cell line. *J. Neurochem.* **74,** 1895–1902

30. Pan, C., Kumar, C., Bohl, S., Klingmueller, U., and Mann, M. (2009) Comparative proteomic phenotyping of cell lines and primary cells to assess preservation of cell type-specific functions. *Mol. Cell. Proteomics* **8,** 443–450

31. Nagaraj, N., Wisniewski, J. R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Paabo, S., and Mann, M. (2011) Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* **7,** 548

32. Beck, M., Schmidt, A., Malmstroem, J., Claassen, M., Ori, A., Szymborska, A., Herzog, F., Rinner, O., Ellenberg, J., and Aebersold, R. (2011) The quantitative proteome of a human cell line. *Mol. Syst. Biol.* **7,** 549

33. Altelaar, A. F., Munoz, J., and Heck, A. J. (2013) Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat. Rev. Genet.* **14,** 35–48

34. Michalski, A. D. E., Hauschild, J. P., Lange, O., Wieghaus, A., Makarov, A., Nagaraj, N., Cox, J., Mann, M., and Horning, S. (2011) Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol. Cell. Proteomics* **10,** M111015

35. Bantscheff, M., Lemeer, S., Savitski, M. M., and Kuster, B. (2012) Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal. Bioanal. Chem.* **404,** 939–965

36. Nagaraj, N., Kulak, N. A., Cox, J., Neuhauser, N., Mayr, K., Hoerning, O., Vorm, O., and Mann, M. (2012) System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. *Mol. Cell. Proteomics* **11,** M111.013722

37. Wisniewski, J. R., Zougman, A., and Mann, M. (2009) Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *J. Proteome Res.* **8,** 5674–5678

38. 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

39. Olsen, J. V., Macek, B., Lange, O., Makarov, A., Horning, S., and Mann, M. (2007) Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods* **4,** 709–712

40. 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

41. Cox, J., Neuhauser, N., Michalski, A., Scheltema, 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

42. Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., and Mann, M. (2014) MaxLFQ allows accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction. *Mol. Cell. Proteomics* M113.031591

43. Schwanhausser, 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

44. Team, R. D. C. (2008) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria

45. Hubner, N. C., Bird, A. W., Cox, J., Splettstoesser, B., Bandilla, P., Poser, I., Hyman, A., and Mann, M. (2010) Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. *J. Cell Biol.* **189,** 739–754

46. Tusher, V. G., Tibshirani, R., and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U.S.A.* **98,** 5116–5121

47. 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., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25,** 25–29

48. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40,** D109–D114

49. Magrane, M., and Consortium, U. (2011) UniProt Knowledgebase: a hub of integrated protein data. *Database* **2011,** bar009

50. Mazzoni, E. O., Mahony, S., Closser, M., Morrison, C. A., Nedelec, S., Williams, D. J., An, D., Gifford, D. K., and Wichterle, H. (2013) Synergistic binding of transcription factors to cell-specific enhancers programs motor neuron identity. *Nat. Neurosci.* **16,** 1219–1227

51. Geiger, T. W. A., Schaab, C., Cox, J., and Mann, M. (2012) Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol. Cell. Proteomics* **11** M111.014050

52. Wisniewski, J. R., Ostasiewicz, P., Dus, K., Zielinska, D. F., Gnad, F., and Mann, M. (2012) Extensive quantitative remodeling of the proteome between normal colon tissue and adenocarcinoma. *Mol. Syst. Biol.* **8,** 611

53. Schmidt, E. R., Pasterkamp, R. J., and van den Berg, L. H. (2009) Axon guidance proteins: novel therapeutic targets for ALS? *Prog. Neurobiol.* **88,** 286–301

54. Philips, T., and Robberecht, W. (2011) Neuroinflammation in amyotrophic lateral sclerosis: role of glial activation in motor neuron disease. *Lancet Neurol.* **10,** 253–263

55. Schaab, C., Geiger, T., Stoehr, G., Cox, J., and Mann, M. (2012) Analysis of high accuracy, quantitative proteomics data in the MaxQB database. *Mol. Cell. Proteomics* **11,** M111.014068

56. Van Den Bosch, L., Van Damme, P., Bogaert, E., and Robberecht, W. (2006) The role of excitotoxicity in the pathogenesis of amyotrophic lateral sclerosis. *Biochim. Biophys. Acta* **1762,** 1068–1082

57. Seeburger, J. L., Tarras, S., Natter, H., and Springer, J. E. (1993) Spinal cord motoneurons express p75NGFR and p145trkB mRNA in amyotrophic lateral sclerosis. *Brain Res.* **621,** 111–115

58. Wu, W. (1996) Potential roles of gene expression change in adult rat spinal motoneurons following axonal injury: a comparison among c-jun, off-

affinity nerve growth factor receptor (LNGFR), and nitric oxide synthase (NOS). *Exp. Neurol.* **141,** 190–200

59. Dupuis, L., Pehar, M., Cassina, P., Rene, F., Castellanos, R., Rouaux, C., Gandelman, M., Dimou, L., Schwab, M. E., Loeffler, J. P., Barbeito, L., and Gonzalez de Aguilar, J. L. (2008) Nogo receptor antagonizes p75NTR-dependent motor neuron death. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 740–745

60. Lu, B., Pang, P. T., and Woo, N. H. (2005) The yin and yang of neurotrophin action. *Nat. Rev. Neurosci.* **6,** 603–614

61. Ash, P. E., Zhang, Y. J., Roberts, C. M., Saldi, T., Hutter, H., Buratti, E., Petrucelli, L., and Link, C. D. (2010) Neurotoxic effects of TDP-43 over-expression in C. elegans. *Hum. Mol. Genet.* **19,** 3206–3218

62. Wils, H., Kleinberger, G., Janssens, J., Pereson, S., Joris, G., Cuijt, I., Smits, V., Ceuterick-de Groote, C., Van Broeckhoven, C., and Kumar-Singh, S. (2010) TDP-43 transgenic mice develop spastic paralysis and neuronal inclusions characteristic of ALS and frontotemporal lobar degeneration. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 3858–3863

63. Jucker, M., and Walker, L. C. (2011) Pathogenic protein seeding in Alzheimer disease and other neurodegenerative disorders. *Ann. Neurol.* **70,** 532–540

Article 2

C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration

Misfolded and aggregated proteins are cardinal features of neurodegenerative disorders. Our collaborators have recently discovered that the hexanucleotide repeat expansion in C9orf72, which marks the most common pathogenic mutation in patients with ALS and FTD, is translated in an ATG start codon independent manner [48]. The resulting poly-dipeptide repeat (pDPR) forms insoluble aggregates in the cytosol. To better understand the pathological mechanism of these insoluble aggregates we investigated their proteome composition. We optimized the sample preparation to break down pDPR associate proteins and were able to determine the corresponding "aggregatome" in primary cortical neurons. As expected, these aggregates are associated with components of the ubiquitin / proteasome machinery like the sequestosome. To identify novel pathological mechanisms, we were specifically interested in proteins that are not associated with protein degradation. Among those, the transport adapter UNC119 showed a complete sequestration into the aggregates. Biological follow up on UNC119 revealed that this sequestration is at least partially responsible for the neurotoxic effects of the aggregates. Moreover, histological analysis of postmortem brain tissue confirmed the sequestration of UNC119 in ALS patients. This article provides a proof of concept for identification of novel disease players by proteomics-based research on toxic aggregates, and makes a significant contribution to the understanding of the general toxic mechanism of protein aggregates in ALS / FTD and neurodegeneration. The paper has been cited 14 times (Google Scholar) within less than a year, illustrating the interest in our findings.

Acta Neuropathologica (2014) [227]

Hornburg D*, May S*, Schludi MH*, Arzberger T, Rentzsch K, Schwenk BM, Grässer FA, Mori K, Kremmer E, Banzhaf-Strathmann J, Mann M, Meissner F, Edbauer D. (*equal contribution)

Contribution: This project was a collaboration with the group of Dieter Edbauer (Ludwig-Maximilian-University Munich). They characterized and provided primary cortical neurons expressing pDPR. I designed a proteomics approach to quantify the composition of the neurotoxic protein aggregates. I developed a sample preparation workflow that allows to breakdown the proteolytic resistant aggregates. Further, I measured and analyzed the data and identified several novel disease relevant proteins such as Unc119.

C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration

Stephanie May · Daniel Hornburg · Martin H. Schludi · Thomas Arzberger · Kristin Rentzsch ·
Benjamin M. Schwenk · Friedrich A. Grässer · Kohji Mori · Elisabeth Kremmer ·
Julia Banzhaf-Strathmann · Matthias Mann · Felix Meissner · Dieter Edbauer

Abstract Hexanucleotide repeat expansion in *C9orf72* is the most common pathogenic mutation in patients with amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration (FTLD). Despite the lack of an ATG start codon, the repeat expansion is translated in all reading frames into dipeptide repeat (DPR) proteins, which form insoluble, ubiquitinated, p62-positive aggregates that are most abundant in the cerebral cortex and cerebellum. To specifically analyze DPR toxicity and aggregation, we expressed DPR proteins from synthetic genes containing a start codon but lacking extensive GGGGCC repeats. Poly-Gly-Ala (GA) formed p62-positive cytoplasmic aggregates, inhibited dendritic arborization and induced apoptosis in primary neurons. Quantitative mass spectrometry analysis to identify poly-GA co-aggregating proteins revealed a significant enrichment of proteins of the ubiquitin–proteasome system. Among the other interacting proteins, we identified the transport factor Unc119, which has been previously linked to neuromuscular and axonal function, as a poly-GA co-aggregating protein. Strikingly, the levels of soluble Unc119 are strongly reduced upon poly-GA expression in neurons, suggesting a loss of function mechanism. Similar to poly-GA expression, Unc119 knockdown inhibits dendritic branching and causes neurotoxicity. Unc119 overexpression partially rescues poly-GA toxicity suggesting that poly-GA expression causes Unc119 loss of function. In *C9orf72* patients, Unc119 is detectable in 9.5 % of GA inclusions in the frontal cortex, but only in 1.6 % of GA inclusions in the cerebellum, an area largely spared of neurodegeneration. A fraction of neurons with Unc119 inclusions shows loss of cytosolic staining. Poly-GA-induced Unc119 loss of function may thereby contribute to selective

S. May, D. Hornburg and M. H. Schludi contributed equally.

Electronic supplementary material The online version of this article (doi:10.1007/s00401-014-1329-4) contains supplementary material, which is available to authorized users.

S. May · M. H. Schludi · T. Arzberger · K. Rentzsch ·
B. M. Schwenk · J. Banzhaf-Strathmann · D. Edbauer (✉)
German Center for Neurodegenerative Diseases (DZNE),
Munich, Schillerstr. 44, 80336 Munich, Germany
e-mail: dieter.edbauer@dzne.de

D. Hornburg · M. Mann · F. Meissner
Max Planck Institute of Biochemistry, Martinsried, Germany

T. Arzberger
Center for Neuropathology and Prion Research, Ludwig-
Maximilians-University Munich, Feodor-Lynen-Str. 23,
81377 Munich, Germany

T. Arzberger
Department of Psychiatry and Psychotherapy, Ludwig-
Maximilians University Munich, Nußbaumstraße 7,
80336 Munich, Germany

F. A. Grässer
Institute of Virology, Saarland University Medical School,
66421 Homburg, Germany

K. Mori · D. Edbauer
Adolf Butenandt Institute, Biochemistry, Ludwig-Maximilians
University Munich, Schillerstr. 44, 80336 Munich, Germany

E. Kremmer
Institute of Molecular Immunology, Helmholtz Zentrum
München, German Research Center for Environmental Health
(GmbH), Marchioninistr. 25, 81377 Munich, Germany

D. Edbauer
Munich Cluster of Systems Neurology (SyNergy), Munich,
Germany

vulnerability of neurons with DPR protein inclusions in the pathogenesis of *C9orf72* FTLD/ALS.

Keywords Neurodegeneration · C9orf72 · FTLD · ALS · Unc119 · Proteomics

Introduction

Amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration (FTLD) are severe neurodegenerative diseases with no effective treatment. Degeneration of the upper and lower motor neurons in ALS leads to progressive paralysis [42]. Depending on the affected regions, FTLD patients suffer from dementia, behavioral abnormalities, language impairment and personality changes [21]. Both diseases have overlapping clinical, neuropathological and genetic features and are often described as extreme ends of a disease spectrum [22].

Recently, a mutation in the non-coding region of the *C9orf72* gene has been identified as the most common genetic cause of both ALS and FTLD [12, 20, 41]. Mutation carriers have a GGGGCC hexanucleotide repeat expansion either in the first intron or the promoter region, depending on the isoform of the *C9orf72* transcript [5]. Patients typically have several hundred or thousand repeats, whereas healthy controls show <33 repeats [5, 51]. *C9orf72* patients exhibit clinical symptoms similar to other FTLD or ALS subtypes, but suffer from an unusually high incidence of psychosis [13].

In addition to the common TDP-43 aggregates in FTLD and ALS, *C9orf72* mutation carriers have abundant star-shaped, TDP-43-negative neuronal cytoplasmic inclusions (NCI) particularly in the cerebellum, hippocampus and frontal neocortex that stain positive for markers of the proteasome system (UPS) such as p62 or ubiquitin [1, 7]. We and others discovered that these TDP-43-negative inclusions contain dipeptide repeat proteins (DPR) that are translated ATG-independent from both sense and antisense transcripts of the *C9orf72* repeat in all reading frames [4, 19, 33, 35, 36, 55]. Repeat translation results in five DPR species, poly-GA, poly-GR, poly-GP, poly-PR and poly-PA. Nearly all TDP-43-negative inclusions contain poly-GA, while the other DPR species co-aggregate to a lesser extent. The translation of the DPR proteins is initiated without an ATG start codon, a phenomenon that was initially discovered in other repeat expansion disorders such as myotonic dystrophy 1 and spinocerebellar ataxia type 8 and was recently also found in fragile X-associated tremor/ataxia syndrome (FXTAS) [48, 54].

Several possible disease mechanisms are discussed (reviewed in [18, 32]). First, DPR protein aggregates, or their precursors, may be toxic through binding or sequestration of cellular proteins. Second, both sense and antisense repeat transcripts accumulate in nuclear RNA foci and may cause the sequestration of specific RNA-binding proteins, which potentially impairs the physiological function of those proteins [15, 26, 43]. Third, *C9orf72* mRNA expression is downregulated in patients with a hexanucleotide repeat expansion, which may indicate a loss of function pathomechanism [12, 20]. Currently, the physiological function of *C9orf72* and the relative importance of the three proposed disease mechanisms are still unclear.

The investigation of aggregation and toxicity of DPR proteins is essential to further elucidate their role in disease progression. Therefore, we developed a primary neuronal cell culture model to test the toxicity and aggregation properties of poly-GA, the most abundant of the five DPR species in patient brain [35]. Our cell-based model reproduces key disease features, including formation of insoluble poly-GA aggregates and co-aggregation with p62. Strikingly, poly-GA expression caused neurotoxicity, suggesting that our cell culture model is a valuable tool to study DPR proteins in vitro. To elucidate the mechanism of GA-mediated neurotoxicity, we analyzed the proteome composition of poly-GA aggregates in our model using mass spectrometry-based proteomics. Recently, we have developed a label-free workflow which allows multiple quantitative comparisons of cellular systems [9, 28] and enables an unbiased analysis of protein aggregates from primary cells. Using this approach, we identified Unc119 as a potential new disease-relevant protein, which is co-aggregating in DPR protein inclusions of *C9orf72* patients.

Materials and methods

Antibodies and reagents

The following antibodies were used: Anti-GFP (mouse N86/8, Neuromab, Davis, CA, USA and rabbit, Clontech, Mountainview, CA, USA), anti-β-actin (Sigma Aldrich, St. Louis, MO, USA), anti-myc (mouse 9E10 and rabbit A-14, Santa Cruz biotechnology, Dallas, TX, USA), anti-p62 for immunoblotting (Cell Signaling, Danvers, MA, USA), anti-p62 for immunofluorescence (MBL, Nagoya, Japan), anti-HA (Sigma Aldrich), anti-V5 (Life technologies, Carlsbad, CA, USA), anti-TDP43 (Sigma Aldrich), anti-phospho-TDP-43 (Ser409/Ser410, rat, clone 1D3) [38], anti-PSD-95 (K28/43, Neuromab), anti-Unc119 (termed Unc119#2, ThermoFisher scientific, Pierce Biotechnology, Rockford, IL, USA), anti-poly-GA [29], anti-PSMC2 (Bethyl laboratories, Montgomery, TX, USA), anti-PSMC4 (Bethyl laboratories), anti-MAP2 (AP-20, Sigma Aldrich), DAPI (Roche Applied Science, Penzberg, Germany). An Unc119 specific antibody (Unc119#1) was raised and affinity

purified against full-length human Unc119 fused to GST in rabbits (Eurogentec, Seraing, Belgium) as described previously [36]. For competition experiments, the diluted Unc119 antibodies were preincubated with native or denatured GST or GST-Unc119 (25 µg/ml) for 2 h at 37 °C. For denaturation, the concentrated GST fusion proteins were heated in 1 % SDS 50 mM Tris pH 8.0 at 95 °C for 5 min. We raised a poly-AP specific monoclonal antibody (clone 14E2 of isotype IgG1) by immunizing rats with synthetic AP_{10} peptide as described previously [29].

DNA constructs and lentivirus production

Synthetic genes for DPR sequences with ATG start codon, reduced GC content and very few remaining GGGGCC repeats were made to order with C-terminal epitope tags (Life technologies, Geneart, Regensburg, Germany). For details and design rational see Fig. S1a. The full sequence information is available in the supplemental methods. Synthetic genes and the original GGGCCG-based poly-GP construct with an ATG start codon were subcloned into pEF6/V5-His vector (Life technologies) or a lentiviral vector driven by human synapsin promoter (FhSynW2). To replace the ATG start codon in the GA_{149}-myc construct with a TAG stop codon we cloned annealed oligonucleotides between an SgrAI site at the 5′ end of the open reading frame and the EcoRI site in the vector. As a negative control GFP from pEGFP-N1 (Clontech) was subcloned into pEF6/V5-His and FhSynW2. The GGGGCC repeat constructs without ATG start codon had been described previously [36]. Rat and human Unc119 cDNA was expressed from a lentiviral vector driven by human ubiquitin promoter containing an N-terminal HA-tag (FUW2-HA). We used shRNA targeting rat Unc119 (GAGAGGCACTACTTTCGAA) or a control targeting firefly luciferase (CGTACGCGGAATACTTCGA) driven by the H1 promoter in the vector FUW coexpressing TagRFP both for transfection and transduction. Lentivirus was produced in HEK293FT cells (Life Technologies) as described previously [17]. The Q_{102}-GFP construct in pCS2 vector was a gift from B. Schmid [44].

Cell culture, immunoblotting and immunofluorescence

HEK293FT cells were transfected using Lipofectamine 2000 according to the manufacturer's instructions. For immunoblotting, cells were harvested in RIPA buffer (137 mM NaCl, 20 mM Tris pH 7.5, 0.1 % SDS, 10 % Glycerol, 1 % Triton X-100, 0.5 % Deoxycholate, 2 mM EDTA) containing protease and phosphatase inhibitor cocktails (Sigma). Cells were lysed on ice for 20 min and centrifuged at low speed to avoid pelleting of the DPR protein aggregates (1,000g for 10 min at 4 °C). The supernatant was mixed with 4 × loading buffer (0.4 M sodium phosphate pH 7.5, 8 % SDS, 40 % glycerol, 10 % 2-mercaptoethanol, bromphenol blue) and incubated at 95 °C for 5 min. Primary hippocampal or cortical neurons were cultured from embryonic day 19 rats and infected with lentiviruses as described previously [17, 47]. Primary cortical neurons infected with indicated lentiviruses were harvested with 2x loading buffer. Samples were run on 12.5 % SDS-PAGE gels or Novex 10–20 % Tris-Tricine gels (Life technologies).

HEK293FT cells and primary neurons were fixed for 10 min with 4 % paraformaldehyde and 4 % sucrose. Primary and secondary antibodies were diluted in GDB buffer (0.1 % gelatine, 0.3 % Triton X-100, 450 mM NaCl, 16 mM sodium phosphate pH 7.4). Confocal images were obtained on a confocal laser scanning LSM710 system (Carl Zeiss, Jena) with a 40 × oil immersion objective. Sholl analysis was performed manually and blinded to the experimental conditions using MetaMorph software as described before [47].

Filter trap assay

To detect DPR aggregates, transfected HEK293FT cells or transduced neurons were harvested with 1 % Triton X-100, 50 mM $MgCl_2$ and 0.2 mg/ml DNase I in PBS. After centrifugation (18,000g for 30 min at 4 °C) the pellet was resuspended in 2 % SDS in 100 mM Tris (pH 7.0). After 1 h incubation at room temperature the homogenates were filtered through a Whatman cellulose acetate membrane with 0.2 µm pore size (Sigma Aldrich).

To detect Unc119 aggregates, brain samples were resuspended in RIPA buffer containing 0.2 mg/ml DNase I. After centrifugation (186,000g for 30 min at 4 °C) the pellet was resuspended in 1 % SDS in 100 mM Tris (pH 7.0) and treated as above.

Cellular assays

Viability of HEK293FT cells and primary neurons was analyzed according to the manufacturer's instructions in 96 well plates: LDH Cytotox Non-Radioactive cytotoxicity assay (Promega), Caspase-glo 3/7 assay (Promega), TUNEL in situ cell death detection TMR red assay (Roche). For the TUNEL assay dead and living cells were counted manually with the Fiji cell counter plugin. At least 400 cells per condition were counted per experiment in a total of three independent experiments. Proteasome activity was measured using the Proteasome-Glo kit according to the manufacturer's instructions (Promega).

qPCR

RT-qPCR of primary cortical neurons was performed as described previously [39]. The following primers were

used for analysis of rat Unc119: GCGCTTTGTTCGATAC-CAGT and TGTTCTTGCTGCTGGGAATG. GAPDH was used as a reference gene: CCGCATCTTCTTGTGCAGT-GCC and AGACTCCACGACATACTCAGCACC.

Immunoprecipitation of poly-GA aggregates

Transduced cortical neurons or transfected HEK293FT cells were harvested in RIPA buffer as described above, additionally adding Benzonase (67 U/ml). Samples were rotated for 30 min at 4 °C prior to centrifugation (1,000*g* for 15 min at 4 °C). 2 % of the input was kept and the rest of the supernatant was added to 50 μl protein G dynabeads (Life Technologies), that were preincubated with 10 μg GFP antibody. After incubation (3 h at 4 °C) the magnetic beads were washed three times (150 mM NaCl, 50 mM Tris pH 7.5, 5 % Glycerol). One-fifth of the bead-mix was denatured in 4× loading buffer (95 °C, 5 min) for western blot analysis and the rest was kept for mass spectrometry (MS) analysis. For co-immunoprecipitations from transfected HEK293FT cells the whole samples were analyzed by western blot.

Sample preparation for MS

The bead-mix was resuspended in 50 μl 8 M Urea, 10 mM Hepes pH 8.0. Protein cysteines were reduced with DTT and alkylated with iodoacetamide (IAA), followed by quenching of IAA with thiourea. Proteins were digested with LysC for 4 h and the bead-mix was centrifuged for 5 min at 16,000*g*. The supernatant was removed and diluted with 4 volumes of 50 mM ammonium bicarbonate. The pellet was resuspended in 1 volume 6 M urea, 2 M thiourea, 10 mM Hepes pH 8.0, 4 volumes 50 mM ammonium bicarbonate and LysC. Trypsin was added to both fractions and the final digest was carried out for 16 h. The resulting peptide mix was desalted on C18 StageTips [40] and analyzed in single shots. Notably, in the supernatant we quantified only 50 proteins (data not shown) whereas over-night digestion of the pellet with LysC and trypsin resulted in over 450 quantifications.

LC–MS/MS

Peptides were separated on a Thermo Scientific EASY-nLC 1000 HPLC system (Thermo Fisher Scientific, Odense, Denmark) via in-house packed columns (75 μm inner diameter, 20 cm length, 1.9 μm C18 particles (Dr. Maisch GmbH, Germany)) in a 100 min gradient from 2 % acetonitrile, 0.5 % formic acid to 80 % acetonitrile, 0.5 % formic acid at 400 nl/min. The column temperature was set to 50 °C. An Orbitrap mass spectrometer [34] (Orbitrap Elite, Thermo Fisher Scientific) was directly coupled to the LC

via nano electrospray source. The Orbitrap Elite was operated in a data-dependent mode. The survey scan range was set from 300 to 1,650 m/z, with a resolution of 120,000. Up to the five most abundant isotope patterns with a charge ≥2 were subjected to collision-induced dissociation fragmentation at a normalized collision energy of 35, an isolation window of 2 Th and a resolution of 15,000 at m/z 200. Data was acquired using the Xcalibur software (Thermo Scientific).

MS data analysis and statistics

To process MS raw files, we employed the MaxQuant software (v 1.4.0.4) [9] and Andromeda search engine [11], against the UniProtKB Rat FASTA database (06/2012) using default settings. Enzyme specificity was set to trypsin allowing cleavage N-terminally to proline and up to 2 miscleavages. Carbamidomethylation was set as fixed modification, acetylation (N-terminus) and methionine oxidation were set as variable modifications. A false discovery rate (FDR) cutoff of 1 % was applied at the peptide and protein level. 'Match between runs', which allows the transfer of peptide identifications in the absence of sequencing, was enabled with a maximum retention time window of 1 min. Protein identification required at least one razor peptide. Data were filtered for common contaminants ($n = 247$). Peptides only identified by site modification were excluded from further analysis. A minimum of two valid quantifications was required in either GA_{149}-GFP or GFP quadruplicates.

For bioinformatic analysis as well as visualization, we used the open PERSEUS environment, which is part of MaxQuant and the R framework (Team, R Development Core, 2008). Imputation of missing values was performed with a normal distribution (width = 0.3; shift = 1.8). For pairwise comparison of proteomes and determination of significant differences in protein abundances, t test statistics were applied with a permutation-based FDR of 2 % and S0 of 2 [50]. For GA-aggregate interacting proteins 1D annotation enrichment on the Welch-test difference using Uniprot Keywords with a Benjamini–Hochberg corrected FDR of 2 % showed a significant enrichment of the annotations for the ubiquitin–proteasome system (gene ontology molecular function: "ubiquitin binding", pfam: "ubiquitin", uniprot keywords: "proteasome") (p value $= 8.7^{-11}$, score = 0.77, 5.7-fold enrichment) [10].

Patient samples

All patient materials were provided by the Neurobiobank Munich, Ludwig-Maximilians-University (LMU) Munich and were collected and distributed according to the guidelines of the local ethical committee.

Clinical data are listed in Table S1. Immunohistochemistry and immunofluorescence stainings were performed as described previously [35]. For competition experiments the Unc119#1 antibody was preincubated with 0.25 µg/µl native GST or GST-Unc119 for 2 h at 37 °C. To compare poly-GA aggregates from patient tissue with aggregates from neuronal culture, non-fixed brain-tissue sample of 1 mm in diameter was smeared between two slides and fixed and stained like cultured neurons. For quantification of Unc119 and GA co-aggregation in the different brain regions three patients were manually analyzed. In each region at least 300 GA aggregates were counted per patient.

Results

Poly-GA forms p62-positive SDS-resistant aggregates in HEK293 cells

To investigate the characteristics of the five different DPR species in cell culture, we generated ATG-initiated epitope-tagged expression constructs for all reading frames of the GGGGCC repeat (Fig. S1a). These synthetic constructs, encoding 149–175 repeats, contain a mixture of alternative codons with reduced GC content to prevent instability observed with repetitive GGGGCC-based constructs in *E. coli,* while allowing for high expression in mammalian cells (Fig. 1). Moreover, changing the original hexanucleotide repeat sequence, but maintaining the DPR protein sequence, allowed us to focus on protein toxicity rather than GGGGCC or CCCCGG RNA toxicity. Unfortunately, gene synthesis for poly-GP constructs repeatedly failed. Thus, we generated an ATG-initiated construct from the endogenous repeat sequence encoding about 80 GP repeats. Importantly, without an ATG start codon GGGGCC repeat constructs did not impair cell viability in HEK293 cells excluding overt RNA toxicity of the utilized constructs (Fig. S1b).

To verify protein expression, we transfected HEK293 cells with the DPR constructs (Fig. S1c). We observed protein products of the expected size for GP_{80}-V5 (21 kDa), but not for GA_{175}-GFP (50 kDa), GFP-GR_{149} (68 kDa), PR_{175}-GFP (71 kDa) and PA_{175}-myc (31 kDa). Similar to immunoblots from patient brains [36], specific bands at the top of the gel for GA_{175}-GFP, GFP-GR_{149}, PR_{175}-GFP and PA_{175}-myc indicate formation of insoluble aggregates for these species (Fig. S1c).

To compare the localization and aggregation of the different DPR species we analyzed transfected HEK293 cells by immunofluorescence (Fig. 1, S2a/b). Strikingly, poly-GA, the most abundant DPR species in patients [35, 36], predominantly formed distinct dot-like or star-shaped

inclusions in the cytosol (Fig. 1) and occasionally in the nucleus (Fig. S2a). In contrast, GFP-GR_{149} showed mainly cytoplasmic staining. PR_{175}-GFP was diffusely localized, both in the cytosol and nucleus. Additionally, GFP-GR_{149} and PR_{175}-GFP expressing cells often showed large dot-like intranuclear inclusions and occasionally smaller cytoplasmic inclusions. In contrast, poly-PA was evenly distributed throughout the nucleus and cytoplasm without apparent aggregation. GP_{80}-V5 was distributed throughout the cytoplasm without forming compact inclusions (Fig. 1).

The poly-GA inclusions were strongly positive for p62 (Fig. 1), suggesting that cytoplasmic poly-GA forms ubiquitinated aggregates similar to the abundant poly-GA inclusions found in *C9orf72* FTLD/ALS [36]. Some GP_{80}-V5 expressing cells also showed increased p62 levels and co-localization with GP_{80}-V5. For the other DPR species no such co-localization was detected. Moreover, immunofluorescence and immunoblotting showed overall increased p62 levels only in GA_{175}-GFP expressing cells (Fig. 1, S1c). In HEK293 none of the constructs induced cell death in an LDH release assay (Fig. S2c).

To confirm aggregation of the DPR proteins, we performed a filter trap assay with HEK293 extracts in the presence of 2 % SDS. Insoluble GA_{175}-GFP and GFP-GR_{149} aggregates were readily detectable on the cellulose acetate filter even upon 125-fold dilution, but no signal was detected for GP_{80}-V5, PR_{175}-GFP and AP_{175}-myc with specific antibodies under these conditions suggesting that they are less aggregation prone and can be solubilized at 2 % SDS in the filter trap assay, but not at 0.1 % SDS in polyacrylamide gels (compare Fig. S1c and S2d).

Taken together, GA_{175}-GFP, GFP-GR_{149} and PR_{175}-GFP DPR proteins formed cytoplasmic or nuclear inclusions in HEK293 cells. Although the number of repeats was different for the individual constructs, these data suggest differential solubility of the five DPR species, since AP_{175}-myc, one of the longest constructs, apparently, remained soluble under these conditions even when omitting the GFP tag. However, we cannot exclude that longer repeats (on average 1,000–2,000) observed in patients may promote aggregation of all DPR species. Importantly, poly-GA mimicked most closely the pathology in patient brain by forming compact p62-positive cytoplasmic inclusions and SDS-resistant aggregates and was therefore used for all subsequent experiments.

Poly-GA forms inclusion in primary hippocampal and cortical neurons

Poly-GA expression in HEK293 cells recapitulates all known features of DPR inclusions seen in *C9orf72* patients, without causing toxicity (Fig. S2c). However, DPR proteins
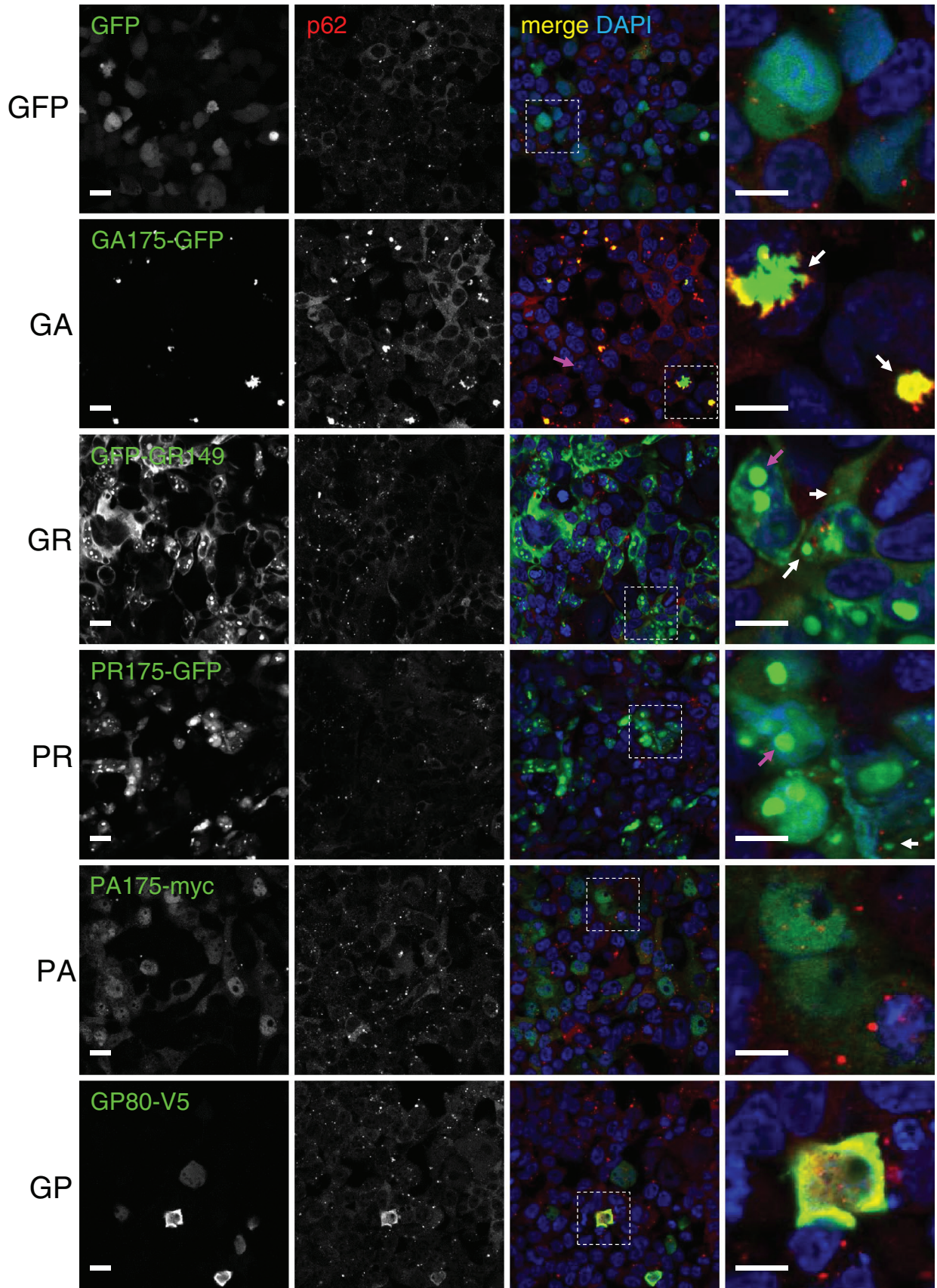
◄ **Fig. 1** DPR species show differential aggregation properties in HEK293 cells. HEK293 cells were transfected with the five different DPR constructs (GA$_{175}$-GFP, GFP-GR$_{149}$, PR$_{175}$-GFP, PA$_{175}$-myc and GP$_{80}$-V5) or GFP as a control and analyzed 2 days later by GFP fluorescence or in case of PA$_{175}$-myc and GP$_{80}$-V5 by immunofluorescence using specific antibodies. DAPI was used as a nuclear marker. Cytoplasmic inclusions (*white arrows*) and nuclear inclusions (*magenta arrows*) are seen for GA$_{175}$-GFP, GFP-GR$_{149}$ and PR$_{175}$-GFP. Many dot-like and star-shaped GA$_{175}$-GFP inclusions co-localize with p62 (*second column from the left*). *Right panels* show close-ups of areas indicated in the *merge column*. Magnifications of intranuclear GA$_{175}$-GFP inclusions are shown in Fig. S2a. Negative control stainings are shown in Fig. S2b. *Scale bar* represents 15 μm for overview and 5 μm for close-up

are almost exclusively expressed in neurons [4, 36] and the *C9orf72* mutation leads to selective degeneration of neurons. Thus, we analyzed the effects of long-term expression of poly-GA in post-mitotic neurons using lentiviral transduction.

Lentiviral expression of GA$_{149}$-GFP in primary rat hippocampal neuron cultures resulted in compact p62-positive poly-GA inclusions (Fig. 2a) similar to the results in HEK293 cells (Fig. 1) and patients [36]. Poly-GA/p62-positive dot-like structures were most common in the cell soma, but were also detectable within dendrites. This finding is reminiscent of the poly-GA-positive dystrophic neurites seen in patient brains [29, 36]. Importantly, the DPR inclusions in transduced neurons and patient neurons showed comparable poly-GA staining intensities suggesting that ATG-driven expression in neurons is a valid model to study DPR toxicity in vitro (Fig. S3). In immunoblots of neuronal extracts all poly-GA protein was retained at the top of the gel indicative of high molecular weight aggregates (Fig. 2b). Consistent with the data in HEK293 cells (Fig. 1, S1c) and patient data [2, 49], p62 levels were strongly increased in poly-GA expressing cells. In contrast, TDP-43 levels were unaffected by GA$_{149}$-myc expression (Fig. 2b) and pathological TDP-43 phosphorylation could not be detected (data not shown). Filter trap analysis further corroborated the formation of SDS-resistant poly-GA aggregates in primary neurons (Fig. 2c).

Poly GA is toxic in primary hippocampal and cortical neurons

Whether DPR proteins contribute to neurodegeneration in *C9orf72* patients is still unclear. In GA$_{149}$-GFP expressing cultures, the neuron density appeared lower although the remaining cells maintained the typical neuronal morphology. However, neurite branching as judged by MAP2 staining appeared less complex (Fig. 2a). Therefore, we quantified dendritic complexity by Sholl analysis, which confirmed reduced branching in GA$_{149}$-myc transfected neurons (Fig. 3a, b).

Furthermore, we quantified neuronal apoptosis in lentivirus transduced cells using several different methods. Compared to controls, GA$_{149}$-myc expressing cortical neurons showed a highly significant 2.0-fold increase in Caspase 3/7 activity (Fig. 3c). Moreover, by analyzing apoptotic DNA fragmentation in primary hippocampal neurons using TUNEL labeling, we detected a highly significant 2.5-fold increase in the number of apoptotic cells (Fig. 3c, compare Fig. S4). Neurotoxicity was also associated with enhanced LDH release in GA$_{149}$-myc expressing cells (Fig. 3e).

To exclude that the synthetic non-GGGGCC repeat sequence encoding GA$_{149}$-myc in our constructs causes RNA-mediated toxicity we replace the ATG start codon with a stop codon (TAG-GA$_{149}$-myc, compare Fig. S1a). Without a start codon we detected no poly-GA expression from the synthetic GA$_{149}$-myc gene upon transduction of primary neurons (Fig. 3d) indicating that this non-GGGGCC construct does not support RAN translation. Importantly, TAG-GA$_{149}$-myc did not impair viability suggesting that the ATG-GA$_{149}$-myc construct causes neurotoxicity due to poly-GA expression and not due to RNA toxicity (Fig. 3e). Therefore, ATG-driven poly-GA expression constructs were used for the remainder of this study.

In summary, poly-GA formed p62-positive inclusions as seen in neurons of patients with *C9orf72* mutation and induced apoptosis in primary cortical and hippocampal neurons, suggesting an important role of poly-GA in the pathogenesis of *C9orf72* FTLD/ALS.

Poly-GA co-aggregates with components of the ubiquitin–proteasome system and the cargo adaptor Unc119

Since DPRs are highly unusual proteins, we wondered if DPR inclusions sequester endogenous proteins and could thereby contribute to disease progression. To this end, we transduced primary cortical neurons with a lentivirus expressing GA$_{149}$-GFP or GFP alone and immunoprecipitated the interacting proteins with anti-GFP in quadruplicates (Fig. S5a). To identify co-aggregating proteins by an unbiased approach we applied label-free quantitative proteomics. By comparing relative protein abundances in GA$_{149}$-GFP and GFP samples we quantified 450 proteins, 20 of which were strongly enriched in poly-GA aggregates (Fig. 4a, Table 1).

Importantly, p62/Sqstm1, a marker protein for DPR inclusions [4, 33, 36, 55], showed strongest enrichment (Fig. 4a), which is consistent with p62 upregulation (Fig. 2b) and p62/GA co-localization (Fig. 2a). Proteasomal subunits (e.g., PSMB6) and other ubiquitin-related proteins (e.g., Ubiquilin 1 and 2) were 5.7-fold enriched in the poly-GA interactome (*p* value = 8.7 × 10^{-11}) (Fig. 4a; Table 1). However, chymotrypsin-like, trypsin-like and caspase-like protease
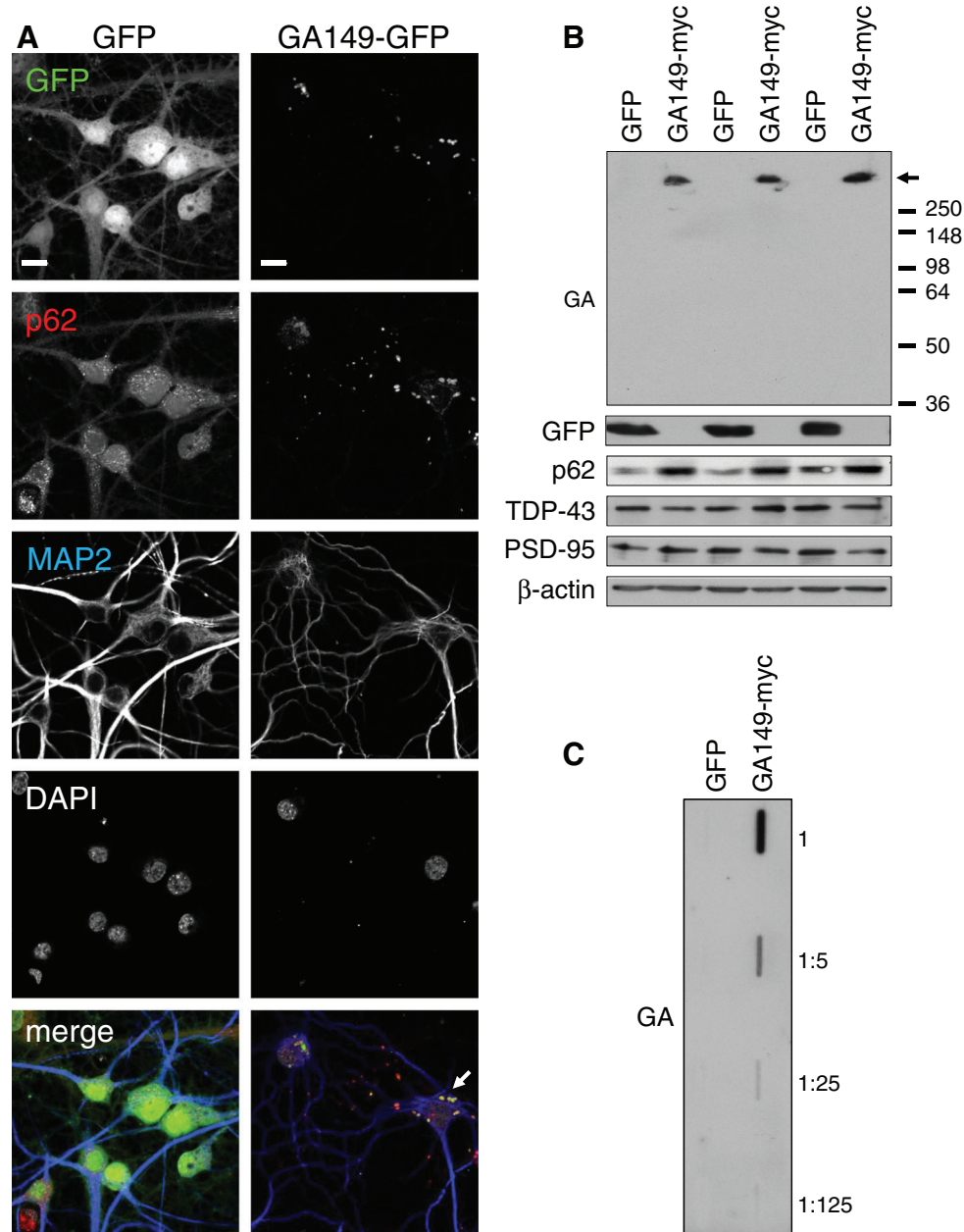
Fig. 2 Poly-GA forms p62-positive aggregates in neurons. **a** Immunofluorescence of primary hippocampal neurons transduced with GA$_{149}$-GFP or GFP control lentivirus at day 6 in vitro for 15 days (DIV6 + 15). Immunostaining for p62 and the dendritic marker protein MAP2. DAPI was used a nuclear marker. Poly-GA forms p62-positive inclusions (*arrow*) in the soma and dendrites. *Scale bar* 15 μm. **b** Immunoblotting of primary cortical neurons transduced with GA$_{149}$-myc or GFP control lentivirus (DIV6 + 17) with the indicated antibodies. Poly-GA aggregates are stuck at the top of the gel (*arrow*). GA$_{149}$-myc induces upregulation of p62, but levels of TDP-43 and the synaptic marker protein PSD-95 are not affected. Three separate transductions are shown. **c** Filter trap assay of primary cortical neurons transduced with GA$_{149}$-myc or GFP (DIV6 + 17). Poly-GA aggregates are detected in the serial dilution of homogenates using anti-GA

activities associated with the proteasome was not impaired in HEK293 cells expressing poly-GA (Fig. S5b). Moreover, the levels of two proteasomal proteins, PSMC2 and PSCM4, were unaffected by poly-GA expression in HEK293 cells and neurons (Fig. S5c/d). TDP-43 was not identified as poly-GA co-aggregating protein which is in line with the lack of

significant co-localization in patients [4, 33, 35, 36, 55]. Interestingly, one of the interaction partners, Unc119, which was 7.5-fold enriched in the GA$_{149}$-GFP immunoprecipitates, was previously identified through severely impaired locomotion in a *C. elegans* mutant and is required for axon development and maintenance [23, 30], which warranted further analysis in the
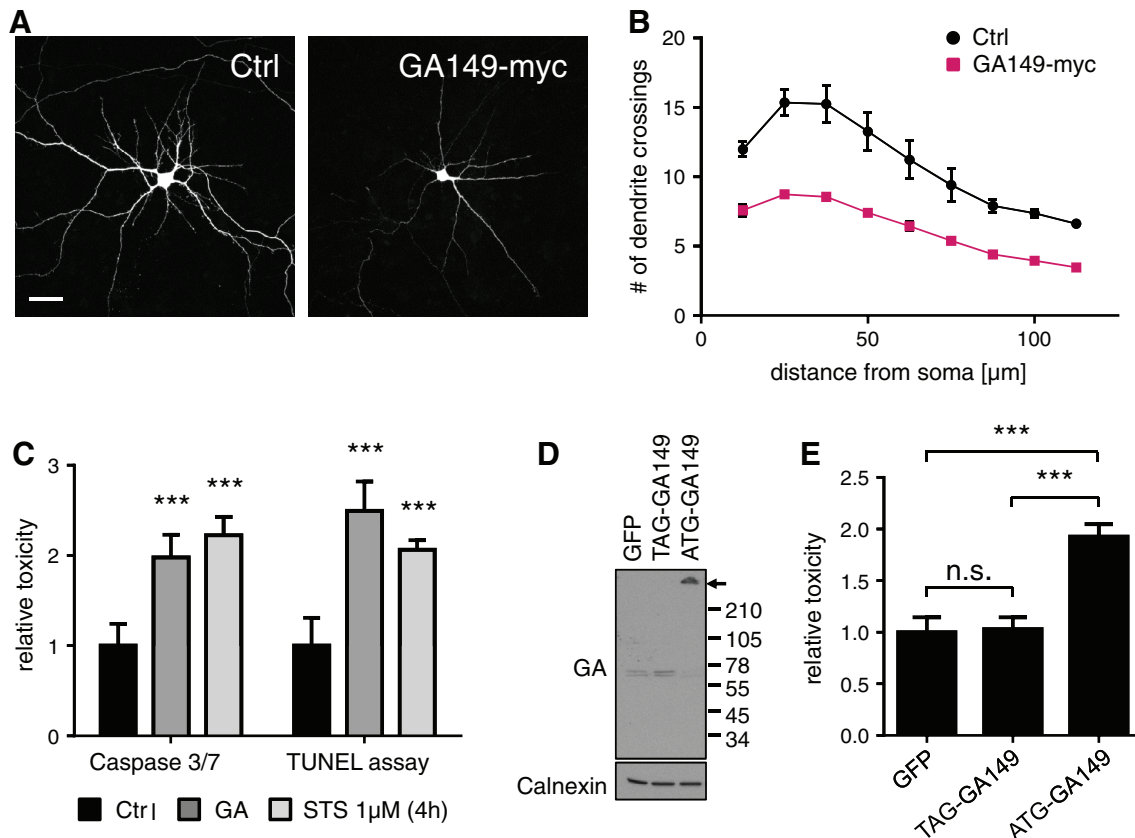
Fig. 3 Poly-GA causes dendrite loss and induces apoptosis in primary neurons. **a** Cortical neurons were co-transfected with empty vector as control (Ctrl) or GA_{149}-myc together with GFP to outline cell morphology (DIV7 + 4). *Scale bars* represent 40 μm. **b** Dendritic complexity was measured using Sholl analysis by manually counting the number of dendrites crossing concentric circles around the soma. Poly-GA expression leads to significant reduction of dendritic branching. $N = 3$ with 40 cells analyzed per condition in each experiment, mean ± SEM. $p < 0.001$ for 12.5 μm radius, $p < 0.0001$ from 25 to 50 μm radius, $p < 0.001$ for 62.5 μm radius, $p < 0.01$ for 75 μm radius and $p < 0.05$ from 87.5 to 112.5 μm radius (two-way ANOVA). **c** Apoptosis in transduced neurons was analyzed using a fluorogenic assay to detect caspase 3/7 activation and a TUNEL assay to detect apoptotic DNA fragmentation (DIV6 + 17). Caspase 3/7 activity was increased 2.0-fold in GA_{149}-myc transduced cortical

neurons. TUNEL-positive apoptotic cells (manually counted using the Fiji cell count plug-in) were increased by 2.5-fold in GA_{149}-myc transduced hippocampal neurons compared to control cells. Representative images of TUNEL stainings are shown in Fig. S2. DIV6 + 17. n = 3 experiments with 6 replicates each; mean ± SD, Student's t test, ***$p < 0.001$. **d** Immunoblots of cortical neurons transduced with GA_{149}-myc constructs with or without start codon (DIV8 + 10). Replacing the ATG start codon in the synthetic GA_{149}-myc gene with a TAG stop codon prevents poly-GA expression and aggregation. *Arrow* indicates top of the gel. **e** LDH release assay detected neurotoxicity of GA_{149}-myc only in the presence of an ATG start codon in transduced cortical neurons (DIV8 + 14). One-way ANOVA with Tukey's post-test. ***$p < 0.001$, n = 3 with six replicates in each experiment

context of ALS. Moreover, Unc119 binds to a myristoylated GAGASA motif of Transducin α (GNAT1), which bears strong resemblance to poly-GA [53]. To confirm that Unc119 interacts and co-aggregates with poly-GA, we co-expressed HA-tagged Unc119 with GA_{175}-GFP in HEK293 cells. This resulted in pronounced co-localization of HA-Unc119 with GA_{175}-GFP inclusions, which is in contrast to the diffuse cytoplasmic localization of HA-Unc119 in GFP expressing cells (Fig. 4b, S6). Co-immunoprecipitation of HA-Unc119 with both GA_{175}-GFP and GA_{149}-myc, but not GFP attests that the interaction is indeed mediated by poly-GA (Fig. S5e). In addition, upon co-expression in HEK293 cells, Unc119 did not co-aggregate with the other DPR species (GFP-GR149,

PR_{175}-GFP, GP_{80}-V5 and PA_{175}-myc) or Q_{102}-GFP, an unrelated aggregating protein [44], supporting a specific interaction of Unc119 and poly-GA (Fig. 4c).

Lentiviral co-expression of Unc119 with GA_{149}-GFP in hippocampal neurons further corroborated the specific co-aggregation of Unc119 with poly-GA (Fig. 5a). Neurons with poly-GA aggregates showed bright Unc119 inclusions, suggesting that a large fraction of cellular Unc119 becomes sequestered in poly-GA inclusions.

In summary, identification of the poly-GA interactome provides proteomic evidence for involvement of the ubiquitin–proteasome system and suggests additional molecular targets of poly-GA toxicity through co-aggregation or sequestration.
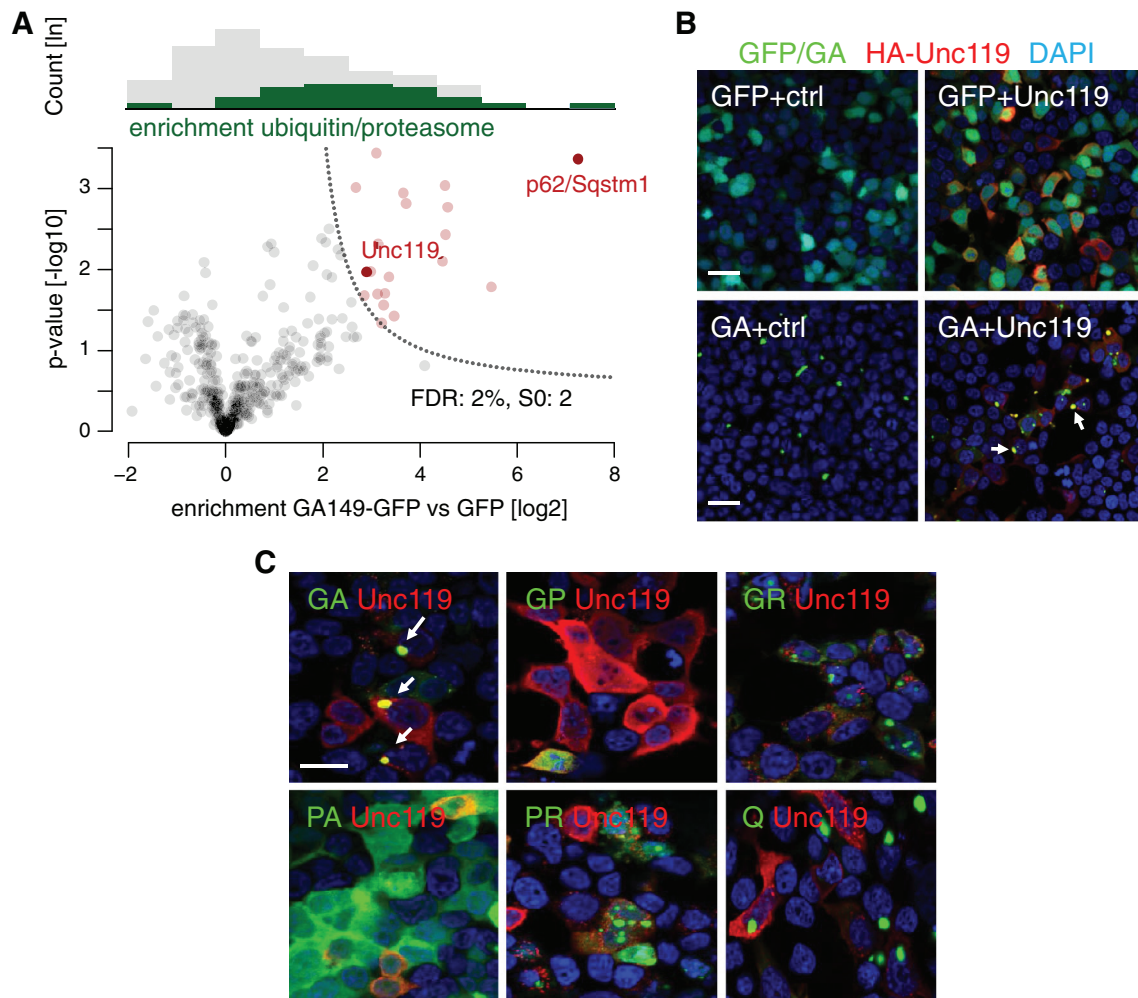
Fig. 4 Unc119 specifically co-aggregates with poly-GA. **a** Quantitative proteomics of GFP immunoprecipitations from primary cortical neurons transduced with GFP or GA_{149}-GFP (DIV6 + 17). p62/Sqstm1 shows highest enrichment and statistical significance. Unc119 was identified by two unique peptides (GGGGTGP-GAEPVPGASNR and LGPLQGK) and one peptide (YQFTPAFLR) shared with its homolog Unc119b. Full protein names are listed in Table 1. *Upper panel* illustrates distribution of quantified protein abundances binned for enrichment factors (*x*-axis below). Enrichment of ubiquitin-related and proteasomal proteins in the poly-GA interactome is highlighted in green. *Lower panel* depicts volcano plot showing poly-GA interacting proteins. False discovery rate (FDR) controlled statistical analysis identified 20 poly-GA interacting proteins compared to control (*red dots*). *Dotted line* depicts threshold for statistical significance. **b** Immunofluorescence of HEK293 cells co-transfected with GFP or GA_{175}-GFP and HA-Unc119 or empty vector control (ctrl). Staining with HA and GFP antibodies and DAPI as nuclear marker. Many GA_{175}-GFP inclusions show co-aggregation of HA-Unc119 (examples marked with *arrows*). Separate channels of these images are shown in Fig. S6. *Scale bar* 30 μm. **c** HEK293 cells were transfected with the five different poly-DPR constructs (GA_{175}-GFP, GFP-GR_{149}, PR_{175}-GFP, PA_{175}-myc and GP_{80}-V5) or Q_{102}-GFP and analyzed using GFP fluorescence and immunostaining of HA-Unc119, PA_{175}-myc and GP_{80}-V5 using specific antibodies 2 days later. HA-Unc119 co-aggregates only with GA_{175}-GFP (*white arrows*). DAPI (*in blue*) was used as a nuclear marker. *Scale bar* 20 μm

Unc119 sequestration contributes to poly-GA toxicity

To analyze how poly-GA inclusions affect endogenous Unc119 we raised a polyclonal antibody against full-length human Unc119 (termed Unc119#1) and tested a commercially available antibody (termed Unc119#2). Both antibodies detected overexpressed rat and human Unc119 (Fig. S7a). To validate both antibodies on endogenous protein, we used RNAi to knockdown Unc119. Lentiviral expression of an

Unc119 specific shRNA in neurons strongly reduced Unc119 mRNA levels compared to control cells (Fig. S7b). Both Unc119 antibodies detected robust knockdown of endogenous Unc119 protein by immunoblotting and immunofluorescence, thus confirming their specificity (Fig. S7c–e).

Although Unc119 was enriched in the poly-GA immunoprecipitation (Fig. 4a), almost no Unc119 could be detected at the regular size (27 kDa) in extracts of GA_{149}-myc expressing neurons by immunoblotting compared to

Table 1 Poly-GA interacting proteins identified by quantitative MS analysis

Gene names	Protein IDs	−log₁₀(p)	log₂(GA/ctrl)	Protein names
Sqstm1	O08623-2; O08623; O08623-3	3.37	7.25	p62/Sequestosome-1
Rad23a; Rad23b	Q4KMA2; Q5XFX7	1.79	5.47	UV excision repair protein RAD23 homolog A and B
Ubqln2; Ubqln4	D4AA63; D4A3P1	2.77	4.57	Ubiquilin 2; Ubiquilin 4
Sdcbp	Q9JI92	2.43	4.52	Syntenin-1
Ubb; Ubc; Uba52; Rps27a; LOC100360645	G3V9Z2; P0CG51; Q63429; F1LML2; Q6P7R7; P62986; Q6PED0; P62982; F1M516	3.04	4.51	Polyubiquitin-B; Ubiquitin; Polyubiquitin-C; Ubiquitin; Ubiquitin-related; Ubiquitin-60S ribosomal protein L40; Ubiquitin; 60S ribosomal protein L40; Ubiquitin-40S ribosomal protein S27a; Ubiquitin; 40S ribosomal protein S27a
Klhdc10	Q5U3Y0; D3ZUK9	2.10	4.46	Kelch domain-containing protein 10
Bag6	Q6MG49; Q6MG49-2	2.81	3.71	Large proline-rich protein BAG6
Psmb6	P28073	2.95	3.65	Proteasome subunit beta type-6
Ubqln1	F1M971; Q9JJP9	1.42	3.47	Ubiquilin 1
Dbn1	Q07266; Q07266-2; C6L8E0	1.91	3.36	Drebrin
Myh10; Myh14	Q9JLT0; G3V9Y1; F1LNF0	1.70	3.28	Myosin-10; Myosin-14
Efhd1; Efhd2	Q4FZY0; D4A9T5	1.56	3.25	EF-hand domain-containing protein D1; EF-hand domain-containing protein D1
Acat2	Q5XI22; F1LS48	1.34	3.21	Acetyl-CoA acetyltransferase, cytosolic
Mlf2	D3ZPN3	2.32	3.13	Myeloid leukemia factor 2
Psmc6	G3V6W6	1.70	3.12	Proteasome (prosome, macropain) 26S subunit, ATPase, 6
Psmb4	D4A640; P34067; G3V8U9	3.44	3.10	Proteasome subunit beta type-4
Psmb5	G3V7Q6; P28075	1.97	2.98	proteasome (prosome, macropain) subunit, beta type, 5
Unc119; Unc119b	Q62885; F1LZN2; D3ZY58	1.97	2.90	Protein unc-119 homolog A and B
Myo5b; Myo5c	F1M111; F1M3R4; P70569	1.68	2.85	Myosin VB/C
Psmd13	B0BN93	3.01	2.68	26S proteasome non-ATPase regulatory subunit 13

Protein enrichment in GA₁₄₉-GFP immunoprecipitates compared to GFP control determined by quantitative mass spectrometry. Listed are the NCBI gene names, the UniProt identifier, logarithmic p value and enrichment factor. Statistical analysis using t test at a false discovery rate of 2 % and S0 of 2 [50]. Proteins with shared peptides are clustered in groups. Graphical representation is depicted in Fig. 4a
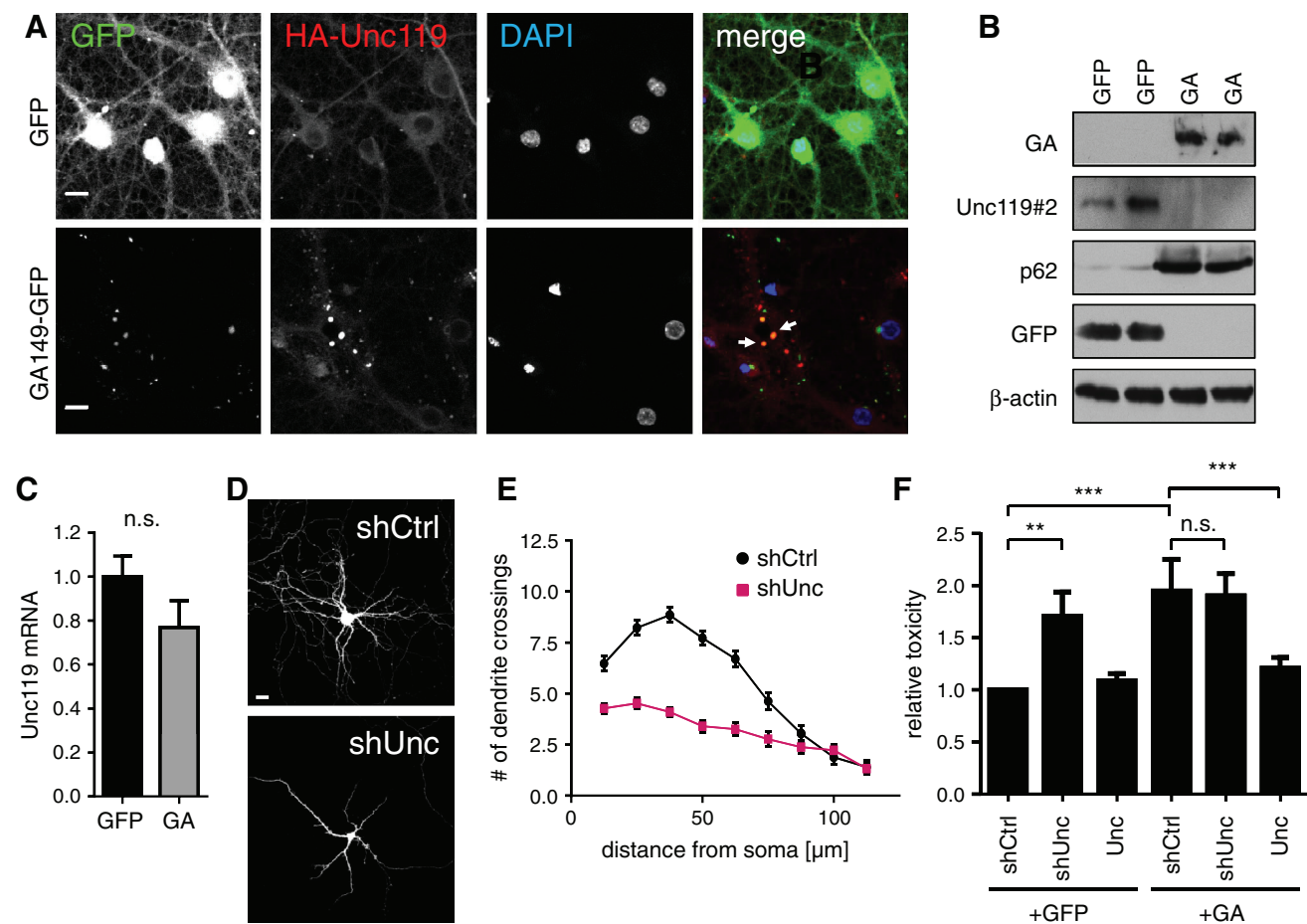
Fig. 5 Unc119 sequestration in neurons contributes to poly-GA toxicity. **a** Immunofluorescence of primary hippocampal neurons co-transduced with HA-Unc119 and either GFP or GA_{149}-GFP (DIV6 + 17). *Arrows* indicate examples of poly-GA inclusions showing co-aggregation of HA-Unc119. *Scale bar* 15 μm. **b** Immunoblot with the indicated antibodies in GFP or GA_{149}-myc transduced cortical neurons shows decreased levels of soluble Unc119 running at 27 kDa. Two separate transductions are shown (DIV6 + 17). **c** qPCR analysis of neurons transduced as in (**b**) shows no significant changes in Unc119 mRNA levels (mean ± SD, Student's *t* test, DIV7 + 10). (**d**, **e**) Hippocampal neurons transfected with shRNA targeting Unc119 (shUnc) or a non-targeting control (shCtrl) together with GFP to outline cell morphology (DIV7 + 5). Dendritic branching was quantified by Sholl analysis. Unc119 knockdown reduced dendrite complexity significantly ($p < 0.0001$ for 12.5–62.5 μm radius and $p < 0.001$ for 75 μm radius, two-way ANOVA, $n = 40$ neurons per condition). *Scale bar* depicts 40 μm. **f** LDH release assay from cortical neurons co-transduced with either GFP or GA_{149}-GFP (GA) together with HA-Unc119 (Unc), shRNA targeting Unc119 (shUnc) or non-targeting shRNA (shCtrl) (DIV6 + 17). Note that Unc119 knockdown causes toxicity in GFP-transduced neurons, but does not increase poly-GA toxicity further. HA-Unc119 expression rescues GA_{149}-GFP toxicity. One-way ANOVA with Tukey's post-test. **$p < 0.01$, ***$p < 0.001$, $n = 3$ with six replicates in each experiment

the GFP expressing control (Fig. 5b). Since the Unc119 mRNA levels remained unchanged (Fig. 5c), this indicates that Unc119 sequestered in poly-GA aggregates becomes insoluble.

To analyze the effect of Unc119 loss of function in neurons we transfected hippocampal neurons with specific shRNAs and analyzed neuron morphology. Unc119 knockdown led to dendrite withering similar to poly-GA expression (Fig. 5d, e). Moreover, compared to a control shRNA, lentiviral Unc119 knockdown induced neuronal death as quantified by increased LDH release (Fig. 5f). While overexpression of HA-Unc119 alone had no effect on cell viability, HA-Unc119 overexpression reduced toxicity in GA_{149}-myc expressing neurons suggesting that Unc119 loss of function contributes to poly-GA toxicity in neurons. In contrast, Unc119 knockdown in GA_{149}-myc expressing neurons did not increase toxicity, which also indicates that Unc119 loss of function is a major source of poly-GA toxicity (Fig. 5f).

In summary, sequestration of Unc119 in poly-GA aggregates may cause Unc119 loss of function and contribute to FTLD/ALS pathogenesis.

Unc119 is a component of DPR inclusions in *C9orf72* patients

Next, we analyzed Unc119 localization in *C9orf72* patients by immunohistochemistry using antibody Unc119#1 to validate co-aggregation with poly-GA found in vitro. In CA3/4 of the hippocampus Unc119 was mainly localized in the cytoplasm (Fig. 6a). Moreover, in all analyzed *C9orf72* FTLD/ALS patients Unc119-positive NCIs were detected, but no Unc119 NCIs were seen in healthy controls (Fig. 6, S8d). In the hippocampus of *C9orf72* patients, Unc119 formed star-shaped NCIs that appeared similar to poly-GA inclusions (Fig. 6b).

Further Unc119 NCIs were detectable in frontal cortex (Fig. 6c, d), occipital cortex (Fig. 6e) and the hippocampal dentate gyrus (Fig. 6f). Importantly, in a fraction of neurons nearly all Unc119 was sequestered into aggregates (Fig. 6b, d, f). Despite abundant DPR pathology only one of the five *C9orf72* cases showed prominent Unc119 NCIs in the cerebellum (Fig. 6g). The second Unc119 antibody (Unc119#2) appeared less sensitive but showed robust NCI pathology in the frontal cortex and in the dentate gyrus (Fig. S8a/b). With both antibodies no Unc119 inclusions were detected in control cases (Fig. 6a, S8c/d).

To further validate antibody specificity we performed competition experiments with GST-Unc119 using immunoblotting (Fig. S9a) and immunohistochemistry (Fig. S9b). Both soluble and inclusion staining were strongly reduced upon preincubation with purified GST-Unc119 further confirming specificity of the Unc119#1 antibody (Fig. S9b). Importantly, this antibody also detected insoluble Unc119 in *C9orf72* patients but not in controls using filter trap (Fig. 6h).

Double immunofluorescence staining with both Unc119 antibodies confirmed co-localization of poly-GA and Unc119 in the cortex and cerebellum of *C9orf72* cases (Fig. 7a, b, S10a). Quantitative analysis in the frontal cortex of three FTLD/ALS patients revealed that Unc119 was present in 9.5 ± 2.7 % of GA inclusions (mean ± standard deviation >300 poly-GA inclusions counted per patient). In contrast, only 0.4–3.3 % of GA inclusions were Unc119 positive in the cerebellum (1.6 ± 1.5 %). In the occipital cortex an intermediate level of co-aggregation was observed (5.8 ± 1.6 %). All Unc119 inclusions were also poly-GA positive suggesting that DPRs drive inclusion formation. Importantly, despite abundant DPR and phospho-TDP-43 pathology in the frontal cortex, there was no co-localization of Unc119 and phospho-TDP-43 within inclusions (Fig. 6c, S10b).

Taken together, Unc119 specifically co-aggregates in poly-GA inclusions in *C9orf72* cases. Notably, Unc119 inclusions were preferentially detected in the frontal cortex, the main region for neurodegeneration in FTLD. Thus, region-specific Unc119 aggregation may contribute to the selective vulnerability of specific neuron populations to *C9orf72* repeat expansion in vivo.

Discussion

Our work establishes a cell culture model for *C9orf72* FTLD/ALS that reproduces core findings in patients and directly links *C9orf72* repeat translation to neurodegeneration. Using quantitative analysis of the poly-GA interactome, we identified a novel co-aggregating protein, Unc119, which has been linked to axon maintenance in *C. elegans* previously [23, 30].

DPR aggregation

Expressing DPR proteins from nearly GGGGCC-free synthetic genes containing ATG start codons allowed us to compare the aggregation properties of the five different DPR species while largely excluding potential secondary effects through RNA toxicity. Previous work with GGGGCC-based expression constructs did not lead to inclusion formation even when a start codon was present [55]. The higher expression levels in our system presumably accelerate disease mechanisms that would normally require gradual build-up of DPR proteins in the brain. In cell culture, the five DPR species displayed remarkably different properties. Only poly-GA expression resulted in compact cytoplasmic inclusions similar to those seen in *C9orf72* mutation brains [33, 36, 55] suggesting that it may be the main driving force for aggregation (Figs. 1, 2). This is in line with the observation that virtually all TDP-43-negative inclusions in *C9orf72* patients contain poly-GA, while antibodies against the other DPR species label only a fraction (10–50 %) of these inclusions [35, 36]. Interestingly, poly-GR and poly-PR formed mainly nuclear inclusions similar to the occasional nuclear DPR inclusions previously identified in patients with poly-GA and p62 antibodies [1, 29, 36]. These two charged DPR species might be actively imported into the nucleus, because a high density of positively charged arginines is also common in classical nuclear localization signals [14]. The discrepancy between aggregation properties observed in patients and our cell culture might be due to the fact, that the synthetic DPR proteins used are much shorter than the several hundred or even thousand repeats found in patients [5, 51].

DPR toxicity

How *C9orf72* repeat expansion leads to neurodegeneration is poorly understood. In fly models RNA toxicity from a 30-mer repeat seems to be the main cause of
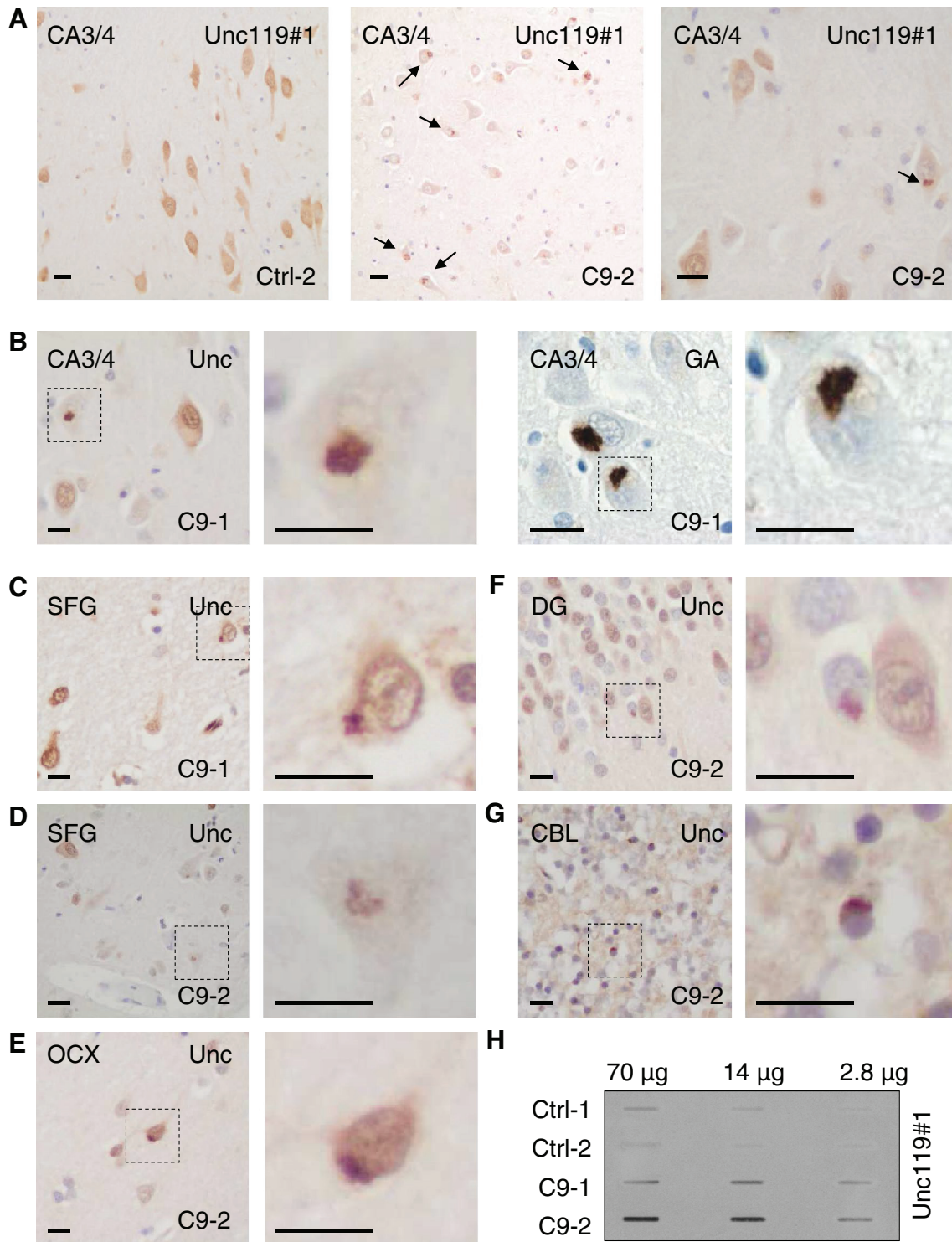
Fig. 6 Unc119 forms neuronal cytoplasmic inclusions in *C9orf72* patients. **a–g** Immunohistochemistry for Unc119 in two *C9orf72* mutation carriers (C9-1 and C9-2) and a control case (Ctrl-2) using antibody Unc119#1. **a** Whereas Unc119 is distributed throughout the cytosol in hippocampal cornu ammonis regions 3/4 (CA3/4) of a control case, a *C9orf72* patient shows neuronal cytoplasmic inclusions. Scale bar represents 20 μm in overviews and 10 μm in close-up. **b** Unc119-positive inclusions have a similar shape as poly-GA inclusions. **c, d** In the superior frontal gyrus (SFG) Unc119-positive cytoplasmic inclusions are detectable in large neurons of mutations carriers. **e–g)** Further Unc119-positive neuronal cytoplasmic inclusions are found in the occipital cortex (OCX) and in the granular cell layers of the dentate gyrus (DG) and the cerebellum (CBL). In various areas, a fraction of cells with Unc119 inclusions shows a clear reduction of cytosolic Unc119 suggesting a redistribution of cytosolic Unc119 into aggregates (close-ups in **b**, **d**, **f**). *Scale bars* represent 10 μm. Counterstains in A-G were done with hemalum. **h** Filter trap assay detects insoluble Unc119 in 1 % SDS in the frontal cortex of mutation carriers, but not in healthy controls
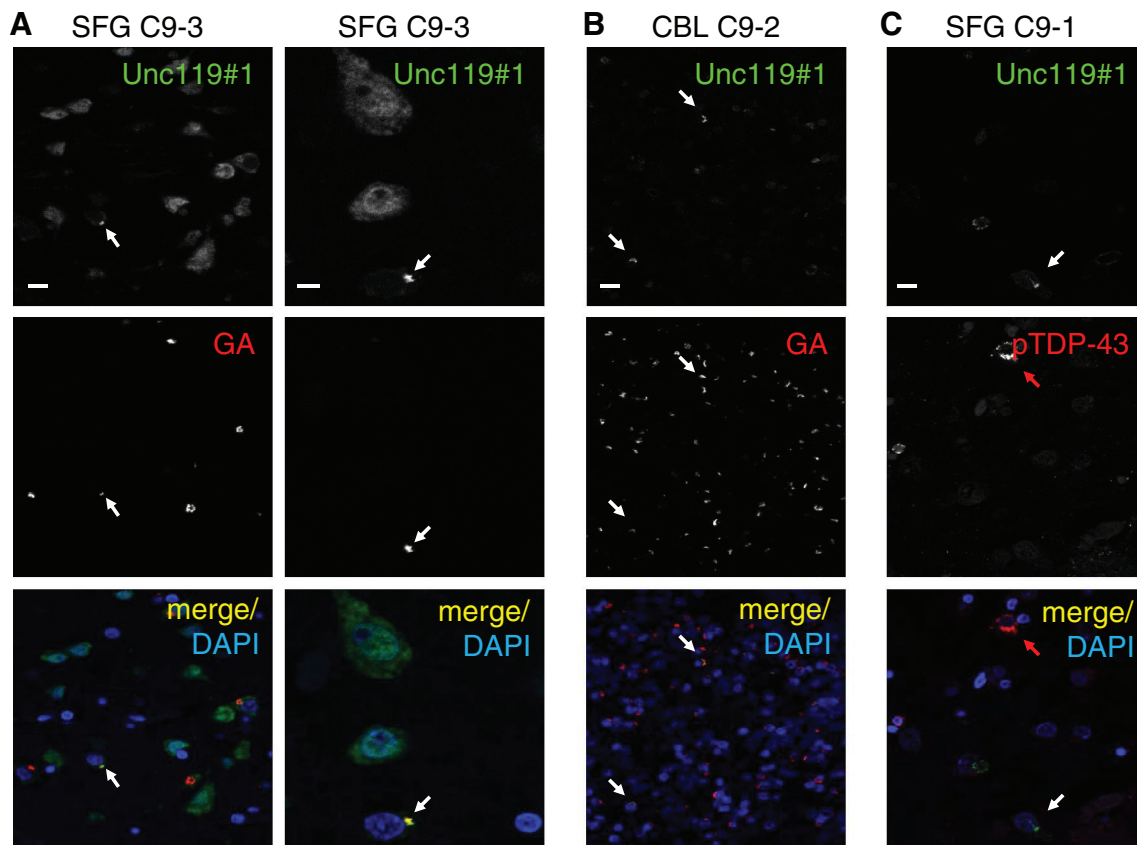
Fig. 7 Unc119 co-aggregates with poly-GA, but not with TDP-43 in patients with *C9orf72* mutation. Double immunofluorescence analysis of Unc119 with poly-GA or phosphorylated TDP-43 (pTDP-43) in *C9orf72* mutation cases C9-1, C9-2 and C9-3. **a** In the superior frontal gyrus (SFG), a subset of poly-GA-positive neuronal cytoplasmic inclusions also contains Unc119. Redistribution of Unc119 compared to GA-negative cells can be seen in a fraction of co-aggregating cells (*white arrows*). **b** In the cerebellar granular cell layer (CBL) abundant cytoplasmic poly-GA inclusions are only rarely positive for Unc119 (*white arrows*). **c** As shown for the superior frontal gyrus, Unc119 (*white arrow*) and pTDP (*red arrow*) are not co-localized in the same cytoplasmic inclusions. *Scale bars* represent 10 μm for overviews and 5 μm for the close-up in the second column

neurodegeneration [52]. Neurons derived from *C9orf72* patients show normal viability, but increased sensitivity to cellular stressors [2, 15, 43]. Zu and colleagues reported combined RNA and protein toxicity for poly-PR and poly-GP in non-neuronal cell culture in the absence of inclusion formation [55]. Despite robust DPR expression in transfected HEK293 cells, we found no evidence for cell death due to protein toxicity in an LDH release assay with the five DPR species. Moreover, the GGGGCC expression constructs without ATG start codon and GGGGCC repeat based poly-GP construct were not toxic, suggesting HEK293 cells are not overtly sensitive to either *C9orf72* repeat RNA or protein toxicity under our conditions. We could not analyze GGGGCC repeat toxicity in neurons, because the repeat seems to block lentiviral packaging. In contrast, caspase activation and DNA fragmentation suggest that p62-positive poly-GA inclusions lead to apoptosis in primary hippocampal and cortical neurons (Figs. 2, 3). Since the synthetic poly-GA gene largely lacks GGGGCC

repeats and requires an ATG start codon to cause toxicity, DPR proteins themselves can cause toxicity in neurons. Due to our construct design these findings, however, do not rule out additional or synergistic effects through GGGGCC repeat-mediated RNA toxicity or *C9orf72* haploinsufficiency in the pathogenesis of *C9orf72* FTLD/ALS. Due to the resemblance of poly-GA aggregates in neurons and patients we focused our study on poly-GA toxicity in neurons. However, it would also be interesting to analyze the effects of other DPR species alone or in combination with poly-GA in neuron culture.

Overexpression models have been invaluable tools to study neurodegenerative diseases but abnormally high levels of the aggregating proteins could also complicate the interpretation [16]. Importantly, lentiviral transduction in our system led to poly-GA aggregates that were comparable in size and poly-GA levels to inclusions from patients suggesting that the observed toxicity of poly-GA in cultured cells is also relevant in vivo (Fig. S3).

Poly-GA interactome

Revealing the interaction profile of poly-GA is an important step to understand the mechanisms leading to the DPR toxicity. Novel instruments, advances of proteomics workflows and new bioinformatics algorithms have greatly increased the accuracy and depth of analysis as well as number of applications for quantitative proteomics [3, 6, 37]. Using GA_{149}-GFP expression, we identified several interacting proteins by affinity purification and quantitative proteomic analysis (Table 1). However, we cannot exclude that additional proteins co-aggregate with the DPR inclusions in *C9orf72* patients. Importantly, the top hit was p62/SQSTM1, an ubiquitin-binding protein that is found in almost all types of intracellular protein aggregates in neurodegenerative diseases including DPR inclusions [1, 25, 36]. This validates our cell culture model and the unique potential of quantitative mass spectrometry to identify disease-relevant protein interactions. Additionally, we found several proteins associated with the ubiquitin proteasome system, but could not detect altered proteasomal expression or activity in poly-GA expressing HEK293 cells or neurons. Interestingly, a Gly/Ala-rich repetitive stretch of about 240 amino acids in EBNA1 was found to block its own proteasomal degradation suggesting that poly-GA may also interfere with the proteasome system [27]. However, in contrast to our findings with poly-GA, the Gly/Ala-rich region in EBNA1 prevents interaction with the proteasome [46], which may be explained by the distinct sequences. EBNA1 only contains GA monomers and dimers and does not form cellular inclusions.

Proteasomal dysfunction has been controversially discussed as a pathomechanism in poly-Q repeat disorders [45]. The poly-G aggregates derived from the CGG repeat expansion in FXTAS are also ubiquitinated [48]. Thus, the ubiquitin proteasome system is clearly linked to repeat expansion diseases although the mechanistic contribution to neurodegeneration remains unclear.

Apart from the ubiquitin–proteasome system, we detected co-localization of poly-GA inclusions with overexpressed and endogenous Unc119, which was among the identified poly-GA interacting proteins (Figs. 4, 5, 7, S10). In the brain, many neurons with Unc119 inclusions show little residual cytosolic Unc119 staining indicating that poly-GA inclusions in *C9orf72* patients lead to partial Unc119 sequestration (Figs. 6, 7). In cultured neurons, poly-GA expression strongly decreases the levels of soluble Unc119 suggesting a possible loss of function component in the disease in brain regions where it aggregates. Interestingly, we only scarcely detect Unc119 inclusions in the cerebellum, an area which shows little neurodegeneration in *C9orf72* patients despite abundant DPR pathology [29]. Unc119 has mainly been studied in the *C. elegans* nervous system and the mammalian retina. Importantly, Unc119 knockout in *C. elegans* almost completely paralyzes the worms and disturbs axonal development and maintenance [23, 30, 31]. Unc119 serves as a trafficking factor for myristoylated proteins, which it specifically binds through a hydrophobic pocket composed of β-sheets [8]. It is intriguing that the binding motif to Transducin α in the retina was mapped to the myristoylated N-terminal GAGASA sequence which strongly supports our interaction data with poly-GA [53]. Apart from this photoreceptor protein the only other known cargos in the nervous system are G_α proteins in the *C. elegans* olfactory system [8]. It will be important to elucidate how Unc119 sequestration affects neuronal function in *C9orf72* patients. We suspect that poly-GA enters and clogs the hydrophobic cavity of Unc119 and thus inhibits transport of so far unidentified myristoylated Unc119 cargos in cortical neurons, which may contribute to neurotoxicity observed upon Unc119 knockdown or poly-GA expression. An Unc119 nonsense mutation was found in a patient with cone rod dystrophy and causes retinal degeneration in mice [24], which is consistent with the toxicity we observed upon Unc119 knockdown in cortical and hippocampal neurons. Importantly, Unc119 overexpression partially rescues poly-GA toxicity in primary neurons, while Unc119 knockdown does not further increase poly-GA toxicity. Together this indicates that Unc119 sequestration is a major cause of poly-GA toxicity.

In conclusion, our data strongly suggest that the unusual translation of the expanded repeats into poly-GA causes neurodegeneration. Co-sequestration of crucial neuronal proteins, such as Unc119, within DPR aggregates may be a novel pathomechanism in *C9orf72* FTLD/ALS further strengthening the importance of DPR aggregates in disease context.

Author contribution SM, KR, BMS, JBS and DE designed, performed and analyzed the cell biological experiments. DH, MM and FM designed, performed and analyzed the proteomic experiments. MHS, SM and KM performed the neuropathological analysis under supervisions of TA. FAG and EK provided reagents. DE and SM wrote the manuscript with input from all coauthors.

References

1. Al-Sarraj S, King A, Troakes C, Smith B, Maekawa S, Bodi I, Rogelj B, Al-Chalabi A, Hortobagyi T, Shaw CE (2011) p62 positive, TDP-43 negative, neuronal cytoplasmic and intranuclear inclusions in the cerebellum and hippocampus define the pathology of C9orf72-linked FTLD and MND/ALS. Acta Neuropathol 122(6):691–702. doi:10.1007/s00401-011-0911-2

2. Almeida S, Gascon E, Tran H, Chou HJ, Gendron TF, Degroot S, Tapper AR, Sellier C, Charlet-Berguerand N, Karydas A, Seeley WW, Boxer AL, Petrucelli L, Miller BL, Gao FB (2013) Modeling key pathological features of frontotemporal dementia with C9ORF72 repeat expansion in iPSC-derived human neurons. Acta Neuropathol 126(3):385–399. doi:10.1007/s00401-013-1149-y

3. Altelaar AF, Munoz J, Heck AJ (2013) Next-generation proteomics: towards an integrative view of proteome dynamics. Nat Rev Genet 14(1):35–48. doi:10.1038/nrg3356

4. Ash PE, Bieniek KF, Gendron TF, Caulfield T, Lin WL, Dejesus-Hernandez M, van Blitterswijk MM, Jansen-West K, Paul JW 3rd, Rademakers R, Boylan KB, Dickson DW, Petrucelli L (2013) Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. Neuron. doi:10.1016/j.neuron.2013.02.004

5. Beck J, Poulter M, Hensman D, Rohrer JD, Mahoney CJ, Adamson G, Campbell T, Uphill J, Borg A, Fratta P, Orrell RW, Malaspina A, Rowe J, Brown J, Hodges J, Sidle K, Polke JM, Houlden H, Schott JM, Fox NC, Rossor MN, Tabrizi SJ, Isaacs AM, Hardy J, Warren JD, Collinge J, Mead S (2013) Large C9orf72 hexanucleotide repeat expansions are seen in multiple neurodegenerative syndromes and are more frequent than expected in the UK population. Am J Hum Genet. doi:10.1016/j.ajhg.2013.01.011

6. Beck M, Schmidt A, Malmstroem J, Claassen M, Ori A, Szymborska A, Herzog F, Rinner O, Ellenberg J, Aebersold R (2011) The quantitative proteome of a human cell line. Mol Syst Biol 7:549. doi:10.1038/msb.2011.82

7. Bigio EH, Weintraub S, Rademakers R, Baker M, Ahmadian SS, Rademaker A, Weitner BB, Mao Q, Lee KH, Mishra M, Ganti RA, Mesulam MM (2012) Frontotemporal lobar degeneration with TDP-43 proteinopathy and chromosome 9p repeat expansion in C9ORF72: clinicopathologic correlation. Neuropathology. doi:10.1111/j.1440-1789.2012.01332.x

8. Constantine R, Zhang H, Gerstner CD, Frederick JM, Baehr W (2012) Uncoordinated (UNC)119: coordinating the trafficking of myristoylated proteins. Vision Res 75:26–32. doi:10.1016/j.visres.2012.08.012

9. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26(12):1367–1372. doi:10.1038/nbt.1511

10. Cox J, Mann M (2012) 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. BMC Bioinform 13(Suppl 16):S12. doi:10.1186/1471-2105-13-S16-S12

11. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res 10(4):1794–1805. doi:10.1021/pr101065j

12. DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, Nicholson AM, Finch NA, Flynn H, Adamson J, Kouri N, Wojtas A, Sengdy P, Hsiung GY, Karydas A, Seeley WW, Josephs KA, Coppola G, Geschwind DH, Wszolek ZK, Feldman H, Knopman DS, Petersen RC, Miller BL, Dickson DW, Boylan KB, Graff-Radford NR, Rademakers R (2011) Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron 72(2):245–256. doi:10.1016/j.neuron.2011.09.011

13. Devenney E, Hornberger M, Irish M, Mioshi E, Burrell J, Tan R, Kiernan MC, Hodges JR (2014) Frontotemporal dementia associated with the C9ORF72 mutation: a unique clinical profile. JAMA Neurol. doi:10.1001/jamaneurol.2013.6002

14. Dingwall C, Robbins J, Dilworth SM, Roberts B, Richardson WD (1988) The nucleoplasmin nuclear location sequence is larger and more complex than that of SV-40 large T antigen. J Cell Biol 107(3):841–849

15. Donnelly CJ, Zhang PW, Pham JT, Heusler AR, Mistry NA, Vidensky S, Daley EL, Poth EM, Hoover B, Fines DM, Maragakis N, Tienari PJ, Petrucelli L, Traynor BJ, Wang J, Rigo F, Bennett CF, Blackshaw S, Sattler R, Rothstein JD (2013) RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. Neuron 80(2):415–428. doi:10.1016/j.neuron.2013.10.015

16. Eisenberg D, Jucker M (2012) The amyloid state of proteins in human diseases. Cell 148(6):1188–1203. doi:10.1016/j.cell.2012.02.022

17. Fleck D, van Bebber F, Colombo A, Galante C, Schwenk BM, Rabe L, Hampel H, Novak B, Kremmer E, Tahirovic S, Edbauer D, Lichtenthaler SF, Schmid B, Willem M, Haass C (2013) Dual cleavage of neuregulin 1 type III by BACE1 and ADAM17 liberates its EGF-like domain and allows paracrine signaling. J Neurosci 33(18):7856–7869. doi:10.1523/JNEUROSCI.3372-12.2013

18. Gendron TF, Belzil VV, Zhang YJ, Petrucelli L (2014) Mechanisms of toxicity in C9FTLD/ALS. Acta Neuropathol 127(3):359–376. doi:10.1007/s00401-013-1237-z

19. Gendron TF, Bieniek KF, Zhang YJ, Jansen-West K, Ash PE, Caulfield T, Daughrity L, Dunmore JH, Castanedes-Casey M, Chew J, Cosio DM, van Blitterswijk M, Lee WC, Rademakers R, Boylan KB, Dickson DW, Petrucelli L (2013) Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. Acta Neuropathol 126(6):829–844. doi:10.1007/s00401-013-1192-8

20. Gijselinck I, Van Langenhove T, van der Zee J, Sleegers K, Philtjens S, Kleinberger G, Janssens J, Bettens K, Van Cauwenberghe C, Pereson S, Engelborghs S, Sieben A, De Jonghe P, Vandenberghe R, Santens P, De Bleecker J, Maes G, Baumer V, Dillen L, Joris G, Cuijt I, Corsmit E, Elinck E, Van Dongen J, Vermeulen S, Van den Broeck M, Vaerenberg C, Mattheijssens M, Peeters K, Robberecht W, Cras P, Martin JJ, De Deyn PP, Cruts M, Van Broeckhoven C (2012) A C9orf72 promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum: a gene identification study. Lancet Neurol 11(1):54–65. doi:10.1016/S1474-4422(11)70261-7

21. Graff-Radford NR, Woodruff BK (2007) Frontotemporal dementia. Semin Neurol 27(01):048–057. doi:10.1055/s-2006-956755

22. Josephs KA, Hodges JR, Snowden JS, Mackenzie IR, Neumann M, Mann DM, Dickson DW (2011) Neuropathological background of phenotypical variability in frontotemporal dementia. Acta Neuropathol 122(2):137–153. doi:10.1007/s00401-011-0839-6

23. Knobel KM, Davis WS, Jorgensen EM, Bastiani MJ (2001) UNC-119 suppresses axon branching in C. elegans. Development 128 (20):4079–4092. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.375.6182

24. Kobayashi A, Higashide T, Hamasaki D, Kubota S, Sakuma H, An W, Fujimaki T, McLaren MJ, Weleber RG, Inana G (2000) HRG4 (UNC119) mutation found in cone-rod dystrophy causes retinal degeneration in a transgenic model. Invest Ophthalmol Vis Sci 41(11):3268–3277

25. Kuusisto E, Kauppinen T, Alafuzoff I (2008) Use of p62/SQSTM1 antibodies for neuropathological diagnosis. Neuropathol Appl Neurobiol 34(2):169–180. doi:10.1111/j.1365-2990.2007.00884.x

26. Lee YB, Chen HJ, Peres JN, Gomez-Deza J, Attig J, Stalekar M, Troakes C, Nishimura AL, Scotter EL, Vance C, Adachi Y, Sardone V, Miller JW, Smith BN, Gallo JM, Ule J, Hirth F, Rogelj B, Houart C, Shaw CE (2013) Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic. Cell Rep 5(5):1178–1186. doi:10.1016/j.celrep.2013.10.049

27. Levitskaya J, Sharipo A, Leonchiks A, Ciechanover A, Masucci MG (1997) Inhibition of ubiquitin/proteasome-dependent protein degradation by the Gly-Ala repeat domain of the Epstein-Barr virus nuclear antigen 1. Proc Natl Acad Sci USA 94(23):12616–12621

28. Luber CA, Cox J, Lauterbach H, Fancke B, Selbach M, Tschopp J, Akira S, Wiegand M, Hochrein H, O'Keeffe M, Mann M (2010) Quantitative proteomics reveals subset-specific viral recognition in dendritic cells. Immunity 32(2):279–289. doi:10.1016/j.immuni.2010.01.013

29. Mackenzie IR, Arzberger T, Kremmer E, Troost D, Lorenzl S, Mori K, Weng SM, Haass C, Kretzschmar HA, Edbauer D, Neumann M (2013) Dipeptide repeat protein pathology in C9ORF72 mutation cases: clinico-pathological correlations. Acta Neuropathol 126(6):859–879. doi:10.1007/s00401-013-1181-y

30. Maduro M, Pilgrim D (1995) Identification and cloning of unc-119, a gene expressed in the Caenorhabditis elegans nervous system. Genetics 141(3):977–988

31. Maduro MF, Gordon M, Jacobs R, Pilgrim DB (2000) The UNC-119 family of neural proteins is functionally conserved between humans Drosophila and C. elegans. J Neurogenet 13(4):191–212

32. Mann D (2014) C9ORF72: grabbing a tiger by the tail. Acta Neuropathol 127(3):311–318. doi:10.1007/s00401-014-1252-8

33. Mann DM, Rollinson S, Robinson A, Bennion Callister J, Thompson JC, Snowden JS, Gendron T, Petrucelli L, Masuda-Suzukake M, Hasegawa M, Davidson Y, Pickering-Brown S (2013) Dipeptide repeat proteins are present in the p62 positive inclusions in patients with frontotemporal lobar degeneration and motor neurone disease associated with expansions in C9ORF72. Acta Neuropathol Commun 1(1):68. doi:10.1186/2051-5960-1-68

34. Michalski A, Damoc E, Lange O, Denisov E, Nolting D, Muller M, Viner R, Schwartz J, Remes P, Belford M, Dunyach JJ, Cox J, Horning S, Mann M, Makarov A (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(3):O111 013698. doi:10.1074/mcp.O111.013698

35. Mori K, Arzberger T, Grasser FA, Gijselinck I, May S, Rentzsch K, Weng SM, Schludi MH, van der Zee J, Cruts M, Van Broeckhoven C, Kremmer E, Kretzschmar HA, Haass C, Edbauer D (2013) Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. Acta Neuropathol 126(6):881–893. doi:10.1007/s00401-013-1189-3

36. Mori K, Weng SM, Arzberger T, May S, Rentzsch K, Kremmer E, Schmid B, Kretzschmar HA, Cruts M, Van Broeckhoven C, Haass C, Edbauer D (2013) The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. Science. doi:10.1126/science.1232927

37. Nagaraj N, Wisniewski JR, Geiger T, Cox J, Kircher M, Kelso J, Paabo S, Mann M (2011) Deep proteome and transcriptome mapping of a human cancer cell line. Mol Syst Biol 7:548. doi:10.1038/msb.2011.81

38. Neumann M, Kwong LK, Lee EB, Kremmer E, Flatley A, Xu Y, Forman MS, Troost D, Kretzschmar HA, Trojanowski JQ, Lee VM (2009) Phosphorylation of S409/410 of TDP-43 is a consistent feature in all sporadic and familial forms of TDP-43 proteinopathies. Acta Neuropathol 117(2):137–149. doi:10.1007/s00401-008-0477-9

39. Orozco D, Tahirovic S, Rentzsch K, Schwenk BM, Haass C, Edbauer D (2012) Loss of fused in sarcoma (FUS) promotes pathological Tau splicing. EMBO Rep 13(8):759–764. doi:10.1038/embor.2012.90

40. Rappsilber J, Mann M, Ishihama Y (2007) Protocol for micropurification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc 2(8):1896–1906. doi:10.1038/nprot.2007.261

41. Renton AE, Majounie E, Waite A, Simon-Sanchez J, Rollinson S, Gibbs JR, Schymick JC, Laaksovirta H, van Swieten JC, Myllykangas L, Kalimo H, Paetau A, Abramzon Y, Remes AM, Kaganovich A, Scholz SW, Duckworth J, Ding J, Harmer DW, Hernandez DG, Johnson JO, Mok K, Ryten M, Trabzuni D, Guerreiro RJ, Orrell RW, Neal J, Murray A, Pearson J, Jansen IE, Sondervan D, Seelaar H, Blake D, Young K, Halliwell N, Callister JB, Toulson G, Richardson A, Gerhard A, Snowden J, Mann D, Neary D, Nalls MA, Peuralinna T, Jansson L, Isoviita VM, Kaivorinne AL, Holtta-Vuori M, Ikonen E, Sulkava R, Benatar M, Wuu J, Chio A, Restagno G, Borghero G, Sabatelli M, Heckerman D, Rogaeva E, Zinman L, Rothstein JD, Sendtner M, Drepper C, Eichler EE, Alkan C, Abdullaev Z, Pack SD, Dutra A, Pak E, Hardy J, Singleton A, Williams NM, Heutink P, Pickering-Brown S, Morris HR, Tienari PJ, Traynor BJ (2011) A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron 72(2):257–268. doi:10.1016/j.neuron.2011.09.010

42. Rowland LP, Shneider NA (2001) Amyotrophic Lateral Sclerosis. N Engl J Med 344(22):1688–1700. doi:10.1056/NEJM200105313442207

43. Sareen D, O'Rourke JG, Meera P, Muhammad AKMG, Grant S, Simpkinson M, Bell S, Carmona S, Ornelas L, Sahabian A, Gendron T, Petrucelli L, Baughn M, Ravits J, Harms MB, Rigo F, Bennett CF, Otis TS, Svendsen CN, Baloh RH (2013) Targeting RNA foci in iPSC-derived motor neurons from ALS patients with a C9ORF72 repeat expansion. Sci Transl Med 5(208):208ra149. doi:10.1126/scitranslmed.3007529

44. Schiffer NW, Broadley SA, Hirschberger T, Tavan P, Kretzschmar HA, Giese A, Haass C, Hartl FU, Schmid B (2007) Identification of anti-prion compounds as efficient inhibitors of polyglutamine protein aggregation in a zebrafish model. J Biol Chem 282(12):9195–9203. doi:10.1074/jbc.M607865200

45. Schipper-Krom S, Juenemann K, Reits EAJ (2012) The ubiquitin-proteasome system in Huntington's disease: are proteasomes impaired, initiators of disease, or coming to the rescue? Biochemistry Research International 2012:12. doi:10.1155/2012/837015

46. Sharipo A, Imreh M, Leonchiks A, Imreh S, Masucci MG (1998) A minimal glycine-alanine repeat prevents the interaction of ubiquitinated I kappaB alpha with the proteasome: a new mechanism for selective inhibition of proteolysis. Nat Med 4(8):939–944

47. Tada T, Simonetta A, Batterton M, Kinoshita M, Edbauer D, Sheng M (2007) Role of Septin cytoskeleton in spine morphogenesis and dendrite development in neurons. Curr Biol 17(20):1752–1758

48. Todd PK, Oh SY, Krans A, He F, Sellier C, Frazer M, Renoux AJ, Chen KC, Scaglione KM, Basrur V, Elenitoba-Johnson K, Vonsattel JP, Louis ED, Sutton MA, Taylor JP, Mills RE, Charlet-Berguerand N, Paulson HL (2013) CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. Neuron 78(3):440–455. doi:10.1016/j.neuron.2013.03.026

49. Troakes C, Maekawa S, Wijesekera L, Rogelj B, Siklos L, Bell C, Smith B, Newhouse S, Vance C, Johnson L, Hortobagyi

T, Shatunov A, Al-Chalabi A, Leigh N, Shaw CE, King A, Al-Sarraj S (2011) An MND/ALS phenotype associated with C9orf72 repeat expansion: abundant p62-positive, TDP-43-negative inclusions in cerebral cortex, hippocampus and cerebellum but without associated cognitive decline. Neuropathology. doi:10.1111/j.1440-1789.2011.01286.x

50. Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci USA 98(9):5116–5121. doi:10.1073/pnas.091062498

51. van der Zee J, Gijselinck I, Dillen L, Van Langenhove T, Theuns J, Engelborghs S, Philtjens S, Vandenbulcke M, Sleegers K, Sieben A, Baumer V, Maes G, Corsmit E, Borroni B, Padovani A, Archetti S, Perneczky R, Diehl-Schmid J, de Mendonca A, Miltenberger-Miltenyi G, Pereira S, Pimentel J, Nacmias B, Bagnoli S, Sorbi S, Graff C, Chiang HH, Westerlund M, Sanchez-Valle R, Llado A, Gelpi E, Santana I, Almeida MR, Santiago B, Frisoni G, Zanetti O, Bonvicini C, Synofzik M, Maetzler W, Vom Hagen JM, Schols L, Heneka MT, Jessen F, Matej R, Parobkova E, Kovacs GG, Strobel T, Sarafov S, Tournev I, Jordanova A, Danek A, Arzberger T, Fabrizi GM, Testi S, Salmon E, Santens P, Martin JJ, Cras P, Vandenberghe R, De Deyn PP, Cruts M, Van Broeckhoven C (2013) A Pan-European study of the C9orf72 repeat associated with FTLD: geographic prevalence, genomic instability and intermediate repeats. Hum Mutat 34(2):363–373. doi:10.1002/humu.22244

52. Xu Z, Poidevin M, Li X, Li Y, Shu L, Nelson DL, Li H, Hales CM, Gearing M, Wingo TS, Jin P (2013) Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. Proc Natl Acad Sci USA 110(19):7778–7783. doi:10.1073/pnas.1219643110

53. Zhang H, Constantine R, Vorobiev S, Chen Y, Seetharaman J, Huang YJ, Xiao R, Montelione GT, Gerstner CD, Davis MW, Inana G, Whitby FG, Jorgensen EM, Hill CP, Tong L, Baehr W (2011) UNC119 is required for G protein trafficking in sensory neurons. Nat Neurosci 14(7):874–880. doi:10.1038/nn.2835

54. Zu T, Gibbens B, Doty NS, Gomes-Pereira M, Huguet A, Stone MD, Margolis J, Peterson M, Markowski TW, Ingram MA, Nan Z, Forster C, Low WC, Schoser B, Somia NV, Clark HB, Schmechel S, Bitterman PB, Gourdon G, Swanson MS, Moseley M, Ranum LP (2011) Non-ATG-initiated translation directed by microsatellite expansions. Proc Natl Acad Sci USA 108(1):260–265. doi:10.1073/pnas.1013343108

55. Zu T, Liu Y, Banez-Coronel M, Reid T, Pletnikova O, Lewis J, Miller TM, Harms MB, Falchook AE, Subramony SH, Ostrow LW, Rothstein JD, Troncoso JC, Ranum LP (2013) RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. Proc Natl Acad Sci USA 110(51):E4968–E4977. doi:10.1073/pnas.1315438110

Article 3

Overexpression of Q-rich prion-like proteins suppresses polyQ cytotoxicity and alters the polyQ interactome

An important question in neurodegeneration is by which mechanisms cells can modify the toxicity of aggregates. While the chaperone, proteasome and autophagy systems represent potent cellular mechanisms to maintain proteostasis, additional pathways might have evolved. In a toxicity screen in yeast, the group of Ulrich Hartl identified proteins that – when overexpressed – can rescue these cells from toxicity of huntingtin aggregates. We investigated whether this rescue correlates with changes in the aggregate composition and based on our results, hypothesized how these proteins can modify the toxicity of aggregates. In this collaborative effort, we chose to focus on the toxicity-modifying transcription factor GTS1. GTS1 itself contains a prion like polyQA domain and we compared the interactome of huntingtin in the presence of overexpressed GTS1 or only the polyQA domain. The reduced htt toxicity significantly correlated with alteration in the aggregate composition. While the association with other prion domain containing proteins decreased, more chaperones associated with the aggregates when either GTS1 or polyQA were overexpressed. *In silico* simulation and imaging data suggested that the structure of aggregates was altered by incorporation of polyQA sequences. In summary, these data provide novel insight into a potential cellular mechanism to detoxify aggregates and suggest a potential therapeutic strategy for neurodegerative diseases.

PNAS (2014) [228]

Ripaud L, Chumakova V, Antonin M, Hastie AR, Pinkert S, Körner R, Ruff KM, Pappu RV, **Hornburg** D, Mann M, Hartl FU, Hipp MS.

Contribution: The Group of Ulrich Hartl (MPI Biochemistry) established a screen, which allowed them to identify genes that suppress polyQ toxicity in yeast. I designed an affinity enrichment protocol to identify polyQ interacting proteins. I further analyzed the proteome composition of the aggregates and statistically assessed the changes in the polyQ interactome on the level of individual proteins and pathways in correlation with the toxicity.

Overexpression of Q-rich prion-like proteins suppresses polyQ cytotoxicity and alters the polyQ interactome

Leslie Ripaud[a,1], Victoria Chumakova[a,1], Matthias Antonin[a,1], Alex R. Hastie[a], Stefan Pinkert[a], Roman Körner[a], Kiersten M. Ruff[b], Rohit V. Pappu[b], Daniel Hornburg[c], Matthias Mann[c], F. Ulrich Hartl[a,2], and Mark S. Hipp[a,2]

Departments of [a]Cellular Biochemistry and [c]Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany; and [b]Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, MO 63130

Expansion of a poly-glutamine (polyQ) repeat in a group of functionally unrelated proteins is the cause of several inherited neurodegenerative disorders, including Huntington's disease. The polyQ length-dependent aggregation and toxicity of these disease proteins can be reproduced in *Saccharomyces cerevisiae*. This system allowed us to screen for genes that when overexpressed reduce the toxic effects of an N-terminal fragment of mutant huntingtin with 103 Q. Surprisingly, among the identified suppressors were three proteins with Q-rich, prion-like domains (PrDs): glycine threonine serine repeat protein (Gts1p), nuclear polyadenylated RNA-binding protein 3, and minichromosome maintenance protein 1. Overexpression of the PrD of Gts1p, containing an imperfect 28 residue glutamine-alanine repeat, was sufficient for suppression of toxicity. Association with this discontinuous polyQ domain did not prevent 103Q aggregation, but altered the physical properties of the aggregates, most likely early in the assembly pathway, as reflected in their increased SDS solubility. Molecular simulations suggested that Gts1p arrests the aggregation of polyQ molecules at the level of nonfibrillar species, acting as a cap that destabilizes intermediates on path to form large fibrils. Quantitative proteomic analysis of polyQ interactors showed that expression of Gts1p reduced the interaction between polyQ and other prion-like proteins, and enhanced the association of molecular chaperones with the aggregates. These findings demonstrate that short, Q-rich peptides are able to shield the interactive surfaces of toxic forms of polyQ proteins and direct them into nontoxic aggregates.

protein aggregation | protein misfolding | neurodegeneration | prion | polyglutamine proteins

Expansion of a poly-glutamine (polyQ) repeat in otherwise unrelated proteins is the cause of several inherited neurological disorders, including Huntington's disease (HD), spino-bulbar muscular atrophy, dentatorubral-pallidoluysian atrophy, and spinocerebellar ataxias 1, 2, 3, 6, 7, and 17 (1). In all these cases, increasing the length of the polyQ repeat over a critical threshold (above 37 Q in HD) (2) results in disease manifestation, with the length of the repeat correlating inversely with age of disease onset (3).

According to the gain-of-toxic function theory, the polyQ expansions increase aggregation propensity and confer to the disease proteins the ability to populate one or more toxic conformations, most likely including various oligomeric and higher-order aggregate states. These aggregate species vary greatly in number of monomeric units, detergent solubility, binding of dyes, and identity and mobility of interacting proteins (4–6). A prominent hypothesis suggests that the pathologic protein aggregates expose novel, highly interactive surfaces that mediate aberrant interactions with other proteins, resulting in their functional impairment and sequestration (7–9). Moreover, the aggregation process is thought to interfere with general protein quality control pathways, including protein folding and the clearance of misfolded proteins (10–13).

The yeast *S. cerevisiae* has been used extensively as a model to explore the basic mechanisms of toxicity mediated by polyQ expansions. The polyQ length dependence of aggregation has been reproduced in yeast upon expression of N-terminal fragments of huntingtin containing the polyQ stretch (*N*-Htt) (14, 15). Interestingly, toxicity, as measured by growth impairment, was found to depend critically on the properties of the sequences flanking the polyQ region (16). Like mammalian prions, yeast prions are able to cause detectable phenotypes without transmitting any genetic material (17), and both *N*-Htt aggregation and toxicity were shown to require the [*RNQ*+] prion (18, 19). Although there are numerous proteins in the yeast proteome that contain Q-rich regions (20), there is no Htt homolog in yeast. Therefore, any toxic effects observed in this model system on *N*-Htt expression can be attributed to a gain-of-toxic function.

Although polyQ aggregation is strongly associated with cell toxicity, the exact nature and the conformational properties of the toxic polyQ species remain elusive. To gain insight into the basic mechanisms of polyQ toxicity, we performed an unbiased yeast genetic screen for suppressors of the growth defect caused by the expression of polyQ expanded *N*-Htt. We identified six genes that reproducibly restored cell growth when overexpressed. Surprisingly, these suppressors include several proteins

with Q-rich sequences, nuclear polyadenylated RNA-binding protein 3 (Nab3p), minichromosome maintenance protein 1 (Mcm1p), and glycine threonine serine repeat protein (Gts1p), which have been suggested to contain candidate prion domains (PrDs) (21). Consistent with results from molecular modeling, quantitative proteomic analysis revealed that these proteins act by shielding the interactive surfaces of polyQ aggregates, thereby reducing their aberrant interactions with endogenous proteins, including multiple bona fide yeast prions, while promoting their association with molecular chaperones.

Materials and Methods

Suppressor Screen. A yeast genomic DNA multicopy library cloned in YEp13 (22, 23) was transformed into YPH499 + pYES2L-103Q-GFP using a high efficiency transformation protocol (24). The cells were plated on synthetic complete (SC) selection medium with galactose to induce 103Q-GFP expression. Of 150,000 transformants, about 200 colonies grew. Each of these colonies was then grown on SC selection medium without 103Q-GFP induction with glucose as the sole carbon source. False positives were eliminated by three successive rounds of screening and positive clones were sequenced (*SI Materials and Methods*).

Interactome Analysis. Yeast cells expressing 103Q-GFP and the suppressor proteins were harvested and lysed. 103Q-GFP was isolated with anti-GFP magnetic beads (Miltenyi). Eluates were desalted and subsequently analyzed by LC coupled to a Q Exactive mass spectrometer (Thermo Fisher) via a nano electrospray source. Data were analyzed using the MaxQuant software (v 1.4.3.19) (25) and label-free algorithms (26). Raw data were searched against the UniProtKB Yeast FASTA database (06/2012). For more details, see *SI Materials and Methods*.

Results

A Yeast Screen for Suppressors of polyQ Toxicity. Previous genetic screens in yeast identified 52 nonessential gene deletions that enhanced the toxicity of a polyQ expanded Htt fragment (27), 28 genes that suppressed toxicity when deleted (28), and 317 genes that suppressed toxicity when overexpressed under control of a *GAL1* promoter (29). However, overexpression of potential suppressors under the strong *GAL1* promoter may lead to variations in copy number of the toxic protein (30) or result in toxicity so that some suppressors may be missed. To overcome these problems, we used a library of genes under the control of their endogenous promoter on a multicopy plasmid, thus resulting in an intermediate expression level (Fig. 1*A*). To this end, we expressed a toxic *N*-Htt-GFP fusion protein with 103 Q, lacking the poly-proline region and containing a N-terminal Flag-tag (103Q-GFP) (Fig. 1*B*) (16), under control of a galactose inducible promoter in yeast positive for the [*RNQ*+] prion. Expression of 103Q-GFP caused a strong growth defect, whereas an otherwise identical protein containing 25 Q (25Q-GFP) caused no growth impairment (Fig. 1*C*) (16, 19).

A yeast genomic DNA multicopy library (22) was transformed into cells harboring the inducible 103Q-GFP construct (Fig. 1*A*). Of 150,000 transformants, ~200 colonies were able to grow on induction of 103Q-GFP. After elimination of colonies that had lost either the expression of 103Q-GFP or the [*RNQ*+] prion, 40 positive clones remained and were analyzed in more detail. Sequencing of the DNA inserts showed that most inserts were selected several times and mapped to 10 different genomic regions, suggesting that the screen was saturated. Four of the isolated inserts encoded genes involved in the metabolism of galactose. These genes were not further analyzed as their expression may have interfered with the expression of 103Q-GFP.

Overview of Screen Results. We identified six different genes that were able to suppress polyQ mediated toxicity (Fig. 1*C*), including the Hsp40 chaperone *SIS1*, the invertase *SUC2*, a gene with unknown function, *SOK1*, and three genes that encode proteins with Q-rich domains: *NAB3, MCM1,* and *GTS1*.
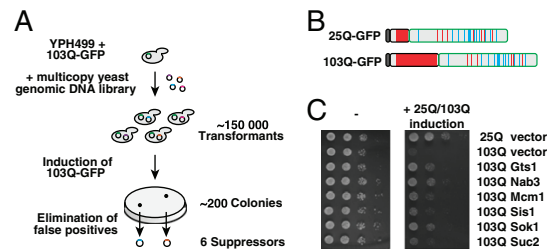

Fig. 1. Identification of suppressors of polyQ toxicity. (*A*) Workflow of overexpression suppressor screen. A yeast genomic DNA library was transformed into cells harboring the inducible 103Q-GFP construct. About 200 colonies were able to grow on induction of 103Q-GFP. False positives were eliminated by three successive rounds of screening. Sequencing of DNA inserts showed that inserts contained 10 different genomic regions. After elimination of inserts that encoded genes involved in the metabolism of galactose, six suppressors of polyQ-mediated toxicity remained. (*B*) Schematic representation of the 25Q/103Q-GFP constructs used in this study. Dark gray indicates the N-terminal Flag-tag, light gray indicates the N terminus of Htt, and green represents the GFP-tag. Glutamines (Q) are shown in red; asparagines (N) in cyan. (*C*) Expression of Gts1p, Nab3p, Mcm1p, Sis1p, Sok1p, and Suc2p suppresses polyQ toxicity. The potential suppressors were preinduced for 48 h before serial dilutions of cells were spotted on galactose medium without doxycycline to coexpress 25Q/103Q-GFP and the suppressors (*Right*) or on glucose plates to control for equal spotting (*Left*).

Sis1p is an essential member of the Hsp40 chaperone family. It functionally cooperates with Hsp70 proteins of the Ssa family and was shown to modulate the aggregation and toxicity of Htt polyQ proteins (14, 19, 31). We recently identified Sis1p as a strong interactor of mutant Htt and as a critical factor in the degradation of misfolded proteins (32). The nonessential protein Sok1p has been shown to suppress mutations in a cAMP-dependent protein kinase, but its exact function is not known (33). *SUC2* has previously been identified as a suppressor of polyQ-mediated toxicity (29) and encodes a nonessential invertase, which hydrolyzes sucrose into glucose and fructose (34). Its mechanism of suppression remains unclear.

The identification of Q-rich, but otherwise unrelated, proteins as suppressors of polyQ-mediated toxicity was intriguing. Notably, a recent bioinformatics screen identified Nab3p, Mcm1p, and Gts1p as potentially prion like (21). Nab3p is a known interactor of expanded polyQ-containing proteins (32, 35) and has an essential role in the termination of RNA polymerase II transcripts (36). Mcm1p is an essential transcription factor involved in pheromone response and regulation of mating type-specific genes (37). Strong overexpression of Mcm1p causes growth arrest (38), but when expressed under control of its endogenous promoter in the multicopy plasmid pRS426, cells grew normally and the growth defect conferred by 103Q-GFP was suppressed (Fig. 1*C*). Gts1p is a nonessential transcription factor that has been implicated in several processes including the stress response, sporulation, and the cell cycle (39, 40). Gts1p was erroneously thought to contain glycine/threonine-serine repeats, but instead contains a Q-rich region including an imperfect glutamine-alanine repeat (QAR) in its C-terminal domain (Fig. 2*A*). Gts1p also contains a Zn-finger motif and an ubiquitin-associated (UBA) domain (40). The Q-rich sequences of the three suppressors differ from those of bona fide prion domains in proteins such as the eukaryotic release factor Sup35p in that they are not enriched in asparagine (N) (Fig. 2*A*).

Q-Rich Domains of Gts1p and Nab3p Are Sufficient to Suppress polyQ Toxicity. The finding that three of the identified suppressors contain uninterrupted polyQ repeats between 9 and 16 residues in length, as well as discontinuous polyQ segments, led us to
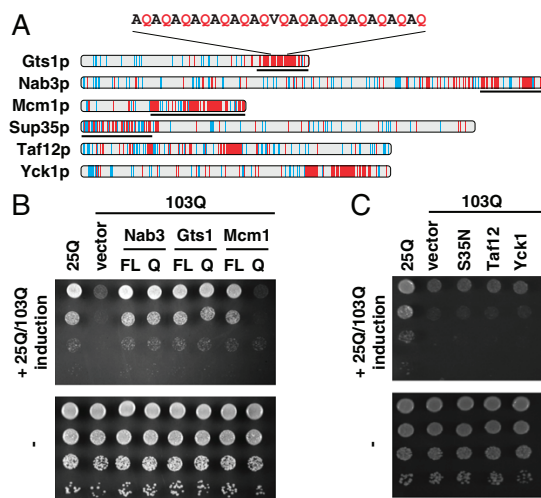
Fig. 2. Effect of Q-rich regions on 103Q-GFP–mediated toxicity. (*A*) Distribution of glutamine residues in Q-rich proteins. Schematic representation of Q and N distribution in the sequences of Q-rich constructs used in this study. N residues, cyan; Q residues, red. Sequences of truncation constructs used in *B* and *C* are indicated by black lines. (*B*) The Q-rich regions of Gts1p and Nab3p alone are sufficient to suppress 103Q-mediated toxicity. Full-length (FL) and truncated Q-rich regions (Q) of *GTS1* and *NAB3* were cloned in pCM190; *MCM1* FL and *MCM1* Q were cloned under the control of the *MCM1* promoter in pRS426. Growth tests were performed as described in Fig. 1C. (*C*) Compositional bias toward glutamines is not sufficient to suppress 103Q-GFP toxicity. The Q-rich proteins Taf12p and Yck1p, as well as the Q-rich N terminus of Sup35p (S35N) were cloned under the control of a tetracycline repressible promoter in pCM190. Growth tests were performed as described in Fig. 1C.

investigate whether the Q-rich domains of Gts1p, Nab3p, and Mcm1p are sufficient for suppression (Fig. 2*A*). Expression of the Q-rich C-terminal regions of Gts1p or Nab3p was sufficient to rescue cell growth to a similar extent as expression of the full-length proteins. In contrast, the corresponding construct of Mcm1p failed to suppress (Fig. 2*B*). Similarly, the Q-rich prion-like proteins Yck1p and Taf12p (21), as well as the prion domain of Sup35p (S35N) (Fig. 2*A*), also failed to suppress 103Q-GFP toxicity, at least at the expression levels tested (Fig. 2*C*). Thus, specific sequence features beyond a compositional bias toward glutamine or the presence of a candidate prion domain must be required for suppression of polyQ toxicity.

The Q-rich domains of Nab3p and Gts1p comprise less than 25% of the sequence of the full-length proteins and contain neither the RNA recognition motif of Nab3p nor the Zn-finger and UBA domain of Gts1p. Thus, it seemed likely that these domains suppress the toxicity of 103Q-GFP by a mechanism unrelated to the biological functions of the full-length proteins. In the remainder of this study, we used Gts1p and its C-terminal region (Gts1Q) as a model to investigate this mechanism.

Gts1p Expression Does Not Change the Prion Status of the Prion Protein Rnq1p. 103Q-GFP toxicity can only be observed in yeast cells when Rnq1p is present in the aggregated prion form [*RNQ*⁺] (19). Curing the [*RNQ*⁺] status will suppress 103Q-GFP–mediated toxicity, and indeed two previously identified loss-of-function suppressors, *rnq1Δ* and *hsp104Δ*, cause the loss of [*RNQ*⁺] (28). Heat shock protein 104 levels have been reported to be altered by Gts1p expression (41), which might influence the Rnq1p prion status and subsequently reduce 103Q-GFP toxicity. To address this possibility, we investigated whether overexpression of Gts1p alters the prion status of Rnq1p, using Rnq1-GFP as a reporter

protein (42). We found that expression of Gts1p neither cured the [*RNQ*⁺] prion nor did it induce [*RNQ*⁺] in cells that had previously been cured by treatment with guanidinium chloride (GdmCl) (Fig. S1*A*). This conclusion is supported further by the observation that 103Q-GFP remains localized in fluorescent foci when Gts1p is overexpressed (Fig. 3*A* and Fig. S1*B*). In contrast, cells cured of the [*RNQ*⁺] prion by GdmCl treatment showed nearly no visible 103Q-GFP foci (Fig. S1*B*) (19).

Gts1p Expression Changes the Aggregation Status of 103Q-GFP. The finding that 103Q-GFP is still present in visible foci on Gts1p overexpression does not exclude the possibility that the presence of Gts1p modifies biophysical properties of the aggregates. The polyQ inclusions that are virtually indistinguishable by light microscopy may nevertheless differ in their physical properties: they may be detergent (SDS) insoluble, containing ordered polyQ fibrils, or detergent soluble, containing structurally amorphous aggregates (15, 43). Using an established filter trap assay for the detection of SDS insoluble aggregates larger than 0.2 μm (44), we found that expression of Gts1p or Gts1Q strongly reduced the amount of SDS insoluble 103Q-GFP present in cells, whereas expression of the N-terminal Gts1p fragment was without effect (Fig. 3*B*). Thus, Gts1p shifts the aggregation properties of the protein from SDS insoluble to SDS soluble aggregate species that still form visible inclusions (Fig. 3*A*), without reducing the amount of 103Q-GFP that can be detected by gel electrophoresis (Fig. 3*C*, input panel).

To demonstrate that Gts1p interacts directly with 103Q-GFP, we added a myc-tag to the N terminus of Gts1p and Gts1Q (Gts1-myc and Gts1Q-myc) to enable detection. Addition of this tag did not change the Gts1p-mediated rescue of 103Q-GFP toxicity (Fig. S2). Although Gts1-myc was diffusely distributed in cells expressing 25Q-GFP (Fig. 3*A*), in 103Q-GFP expressing cells, Gts1-myc was present in bright fluorescent foci colocalizing with 103Q-GFP. Coimmunoprecipitation experiments confirmed the association of Gts1-myc and Gts1Q-myc with 103Q-GFP (Fig. 3*C*), but showed only little association with 25Q-GFP (Fig. S3).
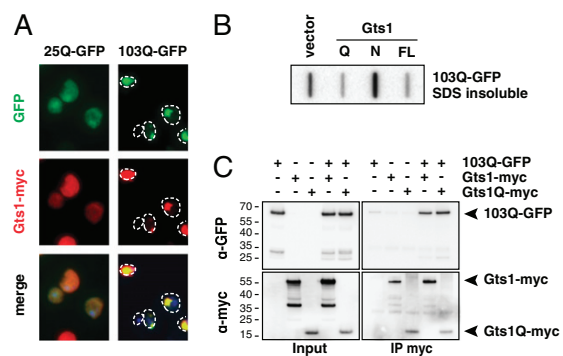

Fig. 3. Gts1p interacts with 103Q-GFP and modulates its aggregation status. (*A*) 103Q-GFP and Gts1-myc colocalize in vivo. YPH499 was transformed with plasmids encoding 25Q/103Q-GFP and Gts1-myc. 25Q/103Q-GFP localization is observed via the fluorescent GFP tag (green). Gts1-myc is observed via indirect fluorescence (red). (*B*) Gts1p expression reduces the amount of SDS insoluble 103Q-GFP. Full-length Gts1p (FL), an N-terminal fragment (N), or a fragment containing the Q-rich C terminus (Q) were preinduced before induction of 103Q-GFP. Cells were lysed and analyzed for SDS insoluble 103Q-GFP by filter retardation assay. (*C*) Gts1-myc and 103Q-GFP physically interact. YPH499 was transformed with plasmids encoding 103Q-GFP and Gts1-myc or Gts1Q-myc. After lysis, anti-myc antibodies were used for immunoprecipitation (IP), followed by SDS/PAGE and Western blotting using anti-GFP or anti-myc antibodies.

Molecular Modeling of Gts1p–polyQ Interactions. Further evidence for direct interactions between the QA-repeat region of Gts1p and the polyQ tract of Htt was obtained by performing atomistic Monte Carlo simulations using the CAMPARI simulation engine (campari.sourceforge.net) and ABSINTH implicit solvation model and forcefield paradigm (45). This approach has recently been used to generate accurate insights regarding the early, nonspecific steps that lead to polyQ aggregation (46, 47) and their modulation by sequences flanking the polyQ tract within Htt (48).

We performed simulations with a polyQ sequence of 35 Q and the imperfect 28 residue QA-region denoted as QAR of Gts1p, present either at a ratio of two molecules of QAR to one molecule of 35Q (a continuous polyQ repeat of 35 Q) or the reverse ratio. The choice of a 35 residue polyQ molecule was made because this allows for simulations that afford reliable and reproducible statistics, and as shown in previous studies, the simulation results are unlikely to vary by increasing or decreasing the polyQ length by five residues vis-à-vis 35Q (46, 47).

Quantifying the statistics for the occurrence of specific distances of separation between the centers of mass of pairs of molecules provided a measure of interaction strength that can be converted into estimates of dissociation constants. We used cumulative distribution functions (CDFs) to quantify the probability that a pair of molecules will assume a specific intermolecular separation at a temperature of 315 K. According to these CDFs, in simulations with a QAR:35Q stoichiometry of 1:2, we find that the probability of heterotypic (QAR-35Q) interactions is equivalent to the probability of homotypic interactions between 35Q molecules (Fig. 4*A*). This result suggests that QAR competes just as effectively for interactions with polyQ as do other polyQ molecules: a finding that is borne out further in simulations with a QAR:35Q stoichiometry of 2:1. Here, we find that QAR molecules prefer interactions with the surfaces of polyQ molecules even in the presence of an additional QAR molecule. Representative conformations drawn from simulations with a QAR:35Q stoichiometry of 1:2 demonstrate the range of interaction modes that are available for QAR to cover the surface of polyQ molecules (Fig. 4*B*). It is worth noting that the ensemble is characterized by considerable structural heterogeneity implying that 35Q and QAR molecules can associate while assuming a range of different conformations. This finding provides a rationale for the observation that Gts1p arrests the aggregation of polyQ molecules at the level of SDS soluble, nonfibrillar species. Taken together, the simulations suggest that QAR is a *trans*-acting cap of polyQ associations that destabilizes intermediates en route to fibrillar species. Interactions of polyQ with QAR presumably generate SDS soluble oligomers with less reactive surfaces. The ability of QA repeats to compete for the surface of polyQ globules within these simulations, in contrast to homotypic interactions between QAR molecules, is supported by the finding that trimers are the dominant species in simulations with a QAR:35Q ratio of 2:1, whereas trimers make up only 50% of the population when the QAR:35Q ratio is 1:2. Importantly, homodimers are prominent when the QAR:35Q ratio is 1:2, whereas heterodimers are preferred over homodimers when the QAR:35Q ratio is 2:1 (Fig. S4). A more extensive analysis of details that emerge from the simulations coupled with in-depth biophysical investigations along the lines of recent efforts (49) is beyond the scope of the current study and will be published elsewhere.

Gts1p Alters the Interaction Properties of 103Q-GFP. PolyQ proteins are widely thought to exert toxicity by mediating aberrant interactions with other proteins, resulting in their functional impairment and sequestration (7, 8, 32, 50). Based on the finding that Gts1p coaggregates with 103Q and consistent with the results of the atomistic simulations, it seemed plausible that Gts1p and Gts1Q suppress polyQ toxicity by interfering with such aberrant interactions. To investigate this possibility, we applied MS-based
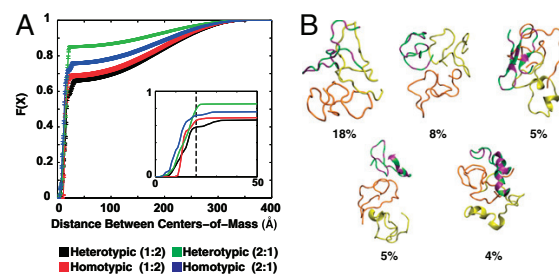

Fig. 4. Molecular modeling of Gts1p-polyQ interactions. (*A*) Cumulative distribution functions quantifying the probabilities for homotypic vs. heterotypic intermolecular interactions. The abscissa denotes the distance (in Angstroms) between the centers of mass of a pair of molecules. The ordinate *F(X)* is the cumulative distribution function (CDF). For a given value of *X*, *F(X)* quantifies the probability that a specific intermolecular separation will be less than or equal to *X*. The black and red curves were generated using statistics from simulations with one molecule containing the imperfect glutamine-alanine repeat of Gts1p (QAR) and two 35Q molecules, respectively. The green and blue curves were generated using statistics from simulations with two QAR molecules and one 35Q molecule. (*Inset*) Zoomed in version of the CDFs to illustrate the quantitative differences and similarities among different curves. It also shows that the distances between the centers of mass are equivalent when each of the *F(X)* curves reaches their respective plateau value. This result is true irrespective of the stoichiometry in the simulations and the type of interactions, heterotypic vs. homotypic, that are being analyzed. (*B*) Representative conformations of trimers formed in simulations with a QAR:35Q ratio of 1:2. The molecules are drawn using backbone traces. The 35Q molecules are shown in orange and yellow, respectively, whereas the QAR molecule is colored using green for the alanine (A)/valine (V) residues and purple for the Q residues. The conformations represent the centers of distinct conformational clusters. Below each representative conformation we show the percent probability of sampling the conformational type within ensembles generated at 315 K. Details of the move sets utilized for the simulations analyzed in *A* and *B* can be found in Table S1.

proteomics. We established a quantitative, label-free workflow that allows an unbiased comparison of multiple samples (25, 26). Using these high accuracy proteomics tools, we quantified the interactome of 103Q-GFP in the absence and presence of Gts1Q expression. Proteins associated with 103Q-GFP were isolated using an anti-GFP antibody coupled to magnetic beads, followed by LC-MS analysis. Interestingly, the group of 103Q-GFP interactors that was most predominantly reduced by expression of Gts1Q ($P = 1.28 \times 10^{-30}$) are known yeast prions or proteins containing PrDs (Fig. 5 and Dataset S1) (21). Sequestration of the PrD containing proteins Sla1p, Sla2p, and Pan1p into aggregates has been suggested to elicit polyQ toxicity by seeding the formation of toxic aggregates (51, 52). Thus, the reduced interaction of these proteins with 103Q-GFP in the presence of Gts1Q can contribute to the suppression of 103Q-GFP toxicity. It is interesting to note that expression of Gts1Q also reduces the interaction of 103Q-GFP with the three prion-like suppressors identified in our screen (Nab3p, Mcm1p, and endogenous Gts1p), thereby also reversing effects possibly due to functional depletion of these proteins, as has been suggested for Nab3p (35).

More than half of the proteins reduced in abundance on 103Q-GFP are nuclear proteins. Many are involved in transcription and RNA metabolic processes, suggesting that interference with these functions contributes to polyQ toxicity. Among the proteins that show an increased interaction with 103Q-GFP in the presence of Gts1Q are several chaperone proteins such as the Hsp70s Ssa3p and Ssa4p, the Hsp90 Hsp82p, and its cochaperone Aha1p. However, there was no change in the interaction of 103Q-GFP with Sis1p, which is the only chaperone identified in our suppressor screen. Comparable results were achieved by overexpression of full-length Gts1p (Fig. S5 and Dataset S1). In this
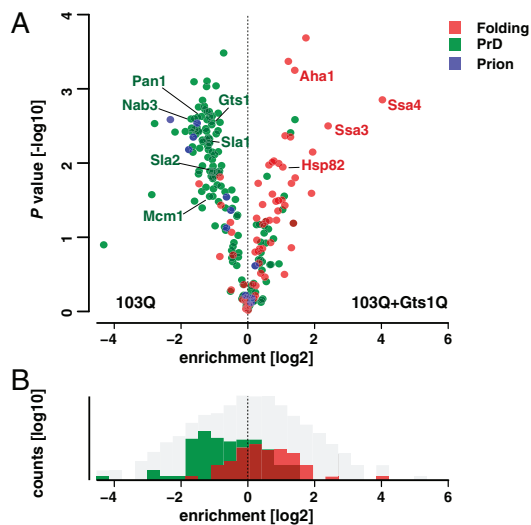
Fig. 5. Gts1Q changes the 103Q-GFP interactome. (*A*) Gts1Q induced alteration of 103Q-GFP interactors. The comparison of *P* value and difference of means from *t* test statistics shows proteins that are enriched, unchanged, or depleted in the comparison of the two interactomes. Interactors involved in protein folding are labeled red, interactors with a predicted prion domain (PrD) are shown in green, and proteins annotated as prions in blue. (*B*) Gts1Q expression increases the number of 103Q-GFP interactors involved in protein folding and decreases the number of interactors with a PrD. The binned distribution of proteins in the pairwise comparison of interactomes shows converse shifts for proteins with a predicted PrD and those involved in protein folding. Interactors involved in protein folding are labeled red, interactors with a predicted prion domain (PrD) are shown in green, and all other identified proteins are shown in gray.

case, the function of Gts1p as transcription factor may alter the composition of the proteome in a manner unrelated to the mechanism described here. The use of Gts1p instead of Gts1Q, for which no tryptic peptides can be detected, allows us to calculate the relative abundances of 103Q-GFP and Gts1p in the interactome. As expected, Gts1p and 103Q-GFP are among the most abundant proteins quantified. The similarity of the normalized MaxLFQ values for Gts1p and 103Q-GFP is consistent with the formation of a 1:1 stoichiometric complex (Fig. S5*C*).

The displacement of functionally critical nuclear components from the polyQ aggregates by Gts1p in combination with the more effective shielding of the aggregates by molecular chaperones can explain how Gts1p suppresses polyQ toxicity. It is also possible that the reduced interaction of the polyQ aggregates with known prions and PrD containing proteins contributes to the suppression of toxicity.

Discussion

Neutralizing the toxic effects of protein aggregates is a rational strategy in the development of treatments for various neurodegenerative disorders, including Huntington's, Parkinson's, and Alzheimer's diseases. In this study, we performed a yeast screen for suppressors of the toxicity of polyQ protein aggregation without making assumptions about the nature of the toxic aggregate species. Among the six suppressors identified were one Hsp40 chaperone, Sis1p, and three proteins that contain C-terminal domains with interrupted polyQ repeats and a compositional bias toward glutamine (Fig. 2*A*). Interestingly, the C-terminal regions of these proteins were previously predicted to contain PrDs (21). For two of the proteins, Gts1p and Nab3p, we demonstrated that expression of the Q-rich C terminus alone is

sufficient to prevent polyQ mediated toxicity (Fig. 2*B*). Thus, these proteins use some shared property of their Q-rich domains to modulate polyQ toxicity, apparently by interacting directly with the polyQ sequence of Htt. Association with the suppressor did not prevent 103Q-GFP aggregation but altered the physical properties of the aggregates. As a result, the association of molecular chaperones of the Hsp70 and Hsp90 systems with the polyQ aggregates was enhanced, and their ability to sequester nuclear proteins was strongly reduced. Protein sequestration into aggregates may lead to an impairment of multiple key cellular pathways and is considered an important mechanism of aggregate toxicity (7, 9, 10, 32, 50, 53–58).

The observed ability of Gts1p and Nab3p to suppress polyQ toxicity was not shared by all proteins containing a PrD (Fig. 2*C*) (30), suggesting that it is based on specific sequence properties of the Q-rich domain and its context with flanking sequences. Our observation that the three PrDs of the suppressors identified in our screen are not enriched for N is in agreement with a recent study that demonstrates that the relative Q and N content of overexpressed PrDs is critical for their ability to suppress polyQ-mediated toxicity (30). Our molecular simulations indicated that the QA stretch might interact with early polyQ intermediates, thereby modulating their surface properties and preventing the formation of large fibrillar aggregates.

Expression of Gts1p markedly reduced the association of yeast prion proteins with the polyQ aggregates. The prion status has been shown to be required for the efficient aggregation of polyQ expansion proteins in yeast, and it has been suggested that the prion aggregates function by seeding toxic polyQ aggregation (18, 19). It seems plausible that Gts1p competes with the yeast prions for interaction with polyQ sequences, thereby shifting the aggregation pathway toward formation of more benign aggregates. Two recent screens suggest that expression of a subset of PrD containing proteins additionally modulates the spatial organization of 103Q-GFP foci (30, 35). At the same time, Gts1p enhances the association of polyQ aggregates with chaperones, which may then cooperate with Gts1p in shielding potentially dangerous surfaces of the polyQ aggregates. Indeed, Gts1Q also reduced the interaction of 103Q-GFP with PrD-containing proteins. A functional depletion of several of these proteins by sequestration into the aggregates has been invoked as a cause of polyQ toxicity (35, 51, 52).

The striking effect of Gts1p in reducing polyQ toxicity may suggest a biological role for proteins with similar QA-repeat sequence motifs in mitigating potentially toxic polyQ interactions, comparable to what has been described for dominant prion mutants (59). Although there is no direct mammalian homolog of Gts1p, the human transcription elongation regulator TCERG1/CA150 also contains a long QA-repeat. Interestingly, TCERG1/CA150 was reported to bind to Htt (60), and overexpression of this transcription factor was shown to protect striatal cells from mutant Htt toxicity in a QA-repeat–dependent manner (61). Expression of peptides with a specific Q-rich region, like the QA-repeat of Gts1p or TCERG1/CA150, might present a strategy to interfere with Htt toxicity. However, regions with QA-repeats may themselves tend to aggregate, as reflected by the aggregation propensity of Gts1p (21, 62) and Cyc8p, another yeast prion with a QA-repeat (63). Indeed, expansion of the QA-repeat of TCERG1/CA150 can lead to an earlier onset of HD (60), presumably by placing a burden on the cellular proteostasis system (10). Nevertheless, controlled expression of peptides containing specific Q-rich domains may be a way to influence Htt toxicity. Future experiments will have to show whether specific interactions between short peptides containing regions similar to the ones discovered in this study can prevent the formation of toxic aggregates and whether this approach can be developed to benefit patients.

1. Orr HT, Zoghbi HY (2007) Trinucleotide repeat disorders. *Annu Rev Neurosci* 30: 575–621.
2. Duyao M, et al. (1993) Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nat Genet* 4(4):387–392.
3. Gusella JF, MacDonald ME (2000) Molecular genetics: Unmasking polyglutamine triggers in neurodegenerative disease. *Nat Rev Neurosci* 1(2):109–115.
4. Stefani M, Dobson CM (2003) Protein aggregation and aggregate toxicity: New insights into protein folding, misfolding diseases and biological evolution. *J Mol Med (Berl)* 81(11):678–699.
5. Winklhofer KF, Tatzelt J, Haass C (2008) The two faces of protein misfolding: Gain- and loss-of-function in neurodegenerative diseases. *EMBO J* 27(2):336–349.
6. Ross CA, Poirier MA (2005) Opinion: What is the role of protein aggregation in neurodegeneration? *Nat Rev Mol Cell Biol* 6(11):891–898.
7. Perez MK, et al. (1998) Recruitment and the role of nuclear localization in polyglutamine-mediated aggregation. *J Cell Biol* 143(6):1457–1470.
8. Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* 75:333–366.
9. Olzscha H, et al. (2011) Amyloid-like aggregates sequester numerous metastable proteins with essential cellular functions. *Cell* 144(1):67–78.
10. Hipp MS, Park SH, Hartl FU (2014) Proteostasis impairment in protein-misfolding and -aggregation diseases. *Trends Cell Biol* 24(9):506–514.
11. Gidalevitz T, Kikis EA, Morimoto RI (2010) A cellular perspective on conformational disease: The role of genetic background and proteostasis networks. *Curr Opin Struct Biol* 20(1):23–32.
12. Powers ET, Morimoto RI, Dillin A, Kelly JW, Balch WE (2009) Biological and chemical approaches to diseases of proteostasis deficiency. *Annu Rev Biochem* 78:959–991.
13. Holmes WM, Klaips CL, Serio TR (2014) Defining the limits: Protein aggregation and toxicity in vivo. *Crit Rev Biochem Mol Biol* 49(4):294–303.
14. Krobitsch S, Lindquist S (2000) Aggregation of huntingtin in yeast varies with the length of the polyglutamine expansion and the expression of chaperone proteins. *Proc Natl Acad Sci USA* 97(4):1589–1594.
15. Muchowski PJ, et al. (2000) Hsp70 and hsp40 chaperones can inhibit self-assembly of polyglutamine proteins into amyloid-like fibrils. *Proc Natl Acad Sci USA* 97(14): 7841–7846.
16. Duennwald ML, Jagadish S, Muchowski PJ, Lindquist S (2006) Flanking sequences profoundly alter polyglutamine toxicity in yeast. *Proc Natl Acad Sci USA* 103(29): 11045–11050.
17. Wickner RB (1994) [URE3] as an altered URE2 protein: Evidence for a prion analog in Saccharomyces cerevisiae. *Science* 264(5158):566–569.
18. Duennwald ML, Jagadish S, Giorgini F, Muchowski PJ, Lindquist S (2006) A network of protein interactions determines polyglutamine toxicity. *Proc Natl Acad Sci USA* 103(29):11051–11056.
19. Meriin AB, et al. (2002) Huntington toxicity in yeast model depends on polyglutamine aggregation mediated by a prion-like protein Rnq1. *J Cell Biol* 157(6):997–1004.
20. Michelitsch MD, Weissman JS (2000) A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions. *Proc Natl Acad Sci USA* 97(22):11910–11915.
21. Alberti S, Halfmann R, King O, Kapila A, Lindquist S (2009) A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell* 137(1):146–158.
22. Nasmyth KA, Tatchell K (1980) The structure of transposable yeast mating type loci. *Cell* 19(3):753–764.
23. Nasmyth KA, Reed SI (1980) Isolation of genes by complementation in yeast: molecular cloning of a cell-cycle gene. *Proc Natl Acad Sci USA* 77(4):2119–2123.
24. Gietz RD, Woods RA (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol* 350:87–96.
25. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26(12):1367–1372.
26. Cox J, et al. (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* 13(9):2513–2526.
27. Willingham S, Outeiro TF, DeVit MJ, Lindquist SL, Muchowski PJ (2003) Yeast genes that enhance the toxicity of a mutant huntingtin fragment or alpha-synuclein. *Science* 302(5651):1769–1772.
28. Giorgini F, Guidetti P, Nguyen Q, Bennett SC, Muchowski PJ (2005) A genomic screen in yeast implicates kynurenine 3-monooxygenase as a therapeutic target for Huntington disease. *Nat Genet* 37(5):526–531.
29. Mason RP, et al. (2013) Glutathione peroxidase activity is neuroprotective in models of Huntington's disease. *Nat Genet* 45(10):1249–1254.
30. Kayatekin C, et al. (2014) Prion-like proteins sequester and suppress the toxicity of huntingtin exon 1. *Proc Natl Acad Sci USA* 111(33):12085–12090.
31. Douglas PM, Summers DW, Ren HY, Cyr DM (2009) Reciprocal efficiency of RNQ1 and polyglutamine detoxification in the cytosol and nucleus. *Mol Biol Cell* 20(19): 4162–4173.
32. Park SH, et al. (2013) PolyQ proteins interfere with nuclear degradation of cytosolic proteins by sequestering the Sis1p chaperone. *Cell* 154(1):134–145.
33. Ward MP, Garrett S (1994) Suppression of a yeast cyclic AMP-dependent protein kinase defect by overexpression of SOK1, a yeast gene exhibiting sequence similarity to a developmentally regulated mouse gene. *Mol Cell Biol* 14(9):5619–5627.
34. Trumbly RJ (1992) Glucose repression in the yeast Saccharomyces cerevisiae. *Mol Microbiol* 6(1):15–21.
35. Wolfe KJ, Ren HY, Trepte P, Cyr DM (2014) Polyglutamine-rich suppressors of huntingtin toxicity act upstream of Hsp70 and Sti1 in spatial quality control of amyloid-like proteins. *PLoS ONE* 9(5):e95914.
36. Conrad NK, et al. (2000) A yeast heterogeneous nuclear ribonucleoprotein complex associated with RNA polymerase II. *Genetics* 154(2):557–571.
37. Elble R, Tye BK (1991) Both activation and repression of a-mating-type-specific genes in yeast require transcription factor Mcm1. *Proc Natl Acad Sci USA* 88(23): 10966–10970.
38. Espinet C, de la Torre MA, Aldea M, Herrero E (1995) An efficient method to isolate yeast genes causing overexpression-mediated growth arrest. *Yeast* 11(1):25–32.
39. Yaguchi S, Mitsui K, Kawabata K, Xu Z, Tsurugi K (1996) The pleiotropic effect of the GTS1 gene product on heat tolerance, sporulation and the life span of Saccharomyces cerevisiae. *Biochem Biophys Res Commun* 218(1):234–237.
40. Bossier P, Goethals P, Rodrigues-Pousada C (1997) Constitutive flocculation in Saccharomyces cerevisiae through overexpression of the GTS1 gene, coding for a 'Glo'-type Zn-finger-containing protein. *Yeast* 13(8):717–725.
41. Yaguchi S, Tsurugi K (2003) Gts1p activates SNF1-dependent derepression of HSP104 and TPS1 in the stationary phase of yeast growth. *J Biol Chem* 278(32):29760–29768.
42. Sondheimer N, Lindquist S (2000) Rnq1: An epigenetic modifier of protein function in yeast. *Mol Cell* 5(1):163–172.
43. Warrick JM, et al. (1999) Suppression of polyglutamine-mediated neurodegeneration in Drosophila by the molecular chaperone HSP70. *Nat Genet* 23(4):425–428.
44. Wanker EE, et al. (1999) Membrane filter assay for detection of amyloid-like polyglutamine-containing protein aggregates. *Methods Enzymol* 309:375–386.
45. Vitalis A, Pappu RV (2009) ABSINTH: A new continuum solvation model for simulations of polypeptides in aqueous solutions. *J Comput Chem* 30(5):673–699.
46. Vitalis A, Wang X, Pappu RV (2008) Atomistic simulations of the effects of polyglutamine chain length and solvent quality on conformational equilibria and spontaneous homodimerization. *J Mol Biol* 384(1):279–297.
47. Vitalis A, Lyle N, Pappu RV (2009) Thermodynamics of beta-sheet formation in polyglutamine. *Biophys J* 97(1):303–311.
48. Williamson TE, Vitalis A, Crick SL, Pappu RV (2010) Modulation of polyglutamine conformations and dimer formation by the N-terminus of huntingtin. *J Mol Biol* 396(5):1295–1309.
49. Crick SL, Ruff KM, Garai K, Frieden C, Pappu RV (2013) Unmasking the roles of N- and C-terminal flanking sequences from exon 1 of huntingtin as modulators of polyglutamine aggregation. *Proc Natl Acad Sci USA* 110(50):20075–20080.
50. Schaffar G, et al. (2004) Cellular toxicity of polyglutamine expansion proteins: Mechanism of transcription factor deactivation. *Mol Cell* 15(1):95–105.
51. Meriin AB, et al. (2003) Aggregation of expanded polyglutamine domain in yeast leads to defects in endocytosis. *Mol Cell Biol* 23(21):7554–7565.
52. Meriin AB, et al. (2007) Endocytosis machinery is involved in aggregation of proteins with expanded polyglutamine domains. *FASEB J* 21(8):1915–1925.
53. Jana NR, Zemskov EA, Wang Gh, Nukina N (2001) Altered proteasomal function due to the expression of polyglutamine-expanded truncated N-terminal huntingtin induces apoptosis by caspase activation through mitochondrial cytochrome c release. *Hum Mol Genet* 10(10):1049–1059.
54. Yamanaka T, et al. (2008) Mutant Huntingtin reduces HSP70 expression through the sequestration of NF-Y transcription factor. *EMBO J* 27(6):827–839.
55. Zhao X, et al. (2012) Sequestration of Sup35 by aggregates of huntingtin fragments causes toxicity of [PSI+] yeast. *J Biol Chem* 287(28):23346–23355.
56. Steffan JS, et al. (2000) The Huntington's disease protein interacts with p53 and CREB-binding protein and represses transcription. *Proc Natl Acad Sci USA* 97(12):6763–6768.
57. Gong H, et al. (2012) Polyglutamine toxicity is controlled by prion composition and gene dosage in yeast. *PLoS Genet* 8(4):e1002634.
58. Kochneva-Pervukhova NV, Alexandrov AI, Ter-Avanesyan MD (2012) Amyloid-mediated sequestration of essential proteins contributes to mutant huntingtin toxicity in yeast. *PLoS ONE* 7(1):e29832.
59. DiSalvo S, Derdowski A, Pezza JA, Serio TR (2011) Dominant prion mutants induce curing through pathways that promote chaperone-mediated disaggregation. *Nat Struct Mol Biol* 18(4):486–492.
60. Holbert S, et al. (2001) The Gln-Ala repeat transcriptional activator CA150 interacts with huntingtin: Neuropathologic and genetic evidence for a role in Huntington's disease pathogenesis. *Proc Natl Acad Sci USA* 98(4):1811–1816.
61. Arango M, et al. (2006) CA150 expression delays striatal cell death in overexpression and knock-in conditions for mutant huntingtin neurotoxicity. *J Neurosci* 26(17): 4649–4659.
62. Sanada M, Kuroda K, Ueda M (2011) GTS1 induction causes derepression of Tup1-Cyc8-repressing genes and chromatin remodeling through the interaction of Gts1p with Cyc8p. *Biosci Biotechnol Biochem* 75(4):740–747.
63. Patel BK, Gavin-Smyth J, Liebman SW (2009) The yeast global transcriptional corepressor protein Cyc8 can propagate as a prion. *Nat Cell Biol* 11(3):344–349.

Article 4

η-Secretase processing of APP inhibits hippocampal neuronal activity

Alzheimer's disease is the most common neurodegenerative disorder. For this disease, the accumulation of extracellular amyloid plaques and intracellular tangles is characteristic. The amyloid plaques are predominantly composed of amyloid-peptides (Aβ), which derive from proteolytic cleavage of the amyloid precursor protein termed APP. For many years, two proteolytic pathways have been held responsible to either promote (β-secretase activity) or suppress (α-secretase activity) generation of toxic Aβ in a competitive manner. Using biochemical and mass spectrometric tools, we identified a third dominant proteolytic pathway for APP, which generates previously unknown neurotoxic fragments, termed Aη. In vivo, membrane bound matrix-metalloproteinases, referred to as η-secretases, cleave APP, releasing long and short Aη fragments after following α-secretase and β- secretase cleavage. These peptides are enriched in dystrophic neurites in an Alzheimer's disease mouse model and human patient brains. Strikingly, when recombinant or synthetic Aη-α was applied on hippocampal slices, long-term potentiation is lowered *ex vivo*. In addition, *in vivo* Ca^{2+} imaging revealed that hippocampal neuronal activity is attenuated by Aη-α. To unravel the exact fragment sequence, we adapted the proteomics workflow and employed the latest generation of hybrid quadrupole Orbitrap instrumentation. We combined various proteases of different specificities with partial digestion. This approach allowed us to boost the evidence for the Aη fragments several fold compared to conventional sample preparation and to map the novel cleavage site 'Aη'. These data highlight a third, so for unknown pathway in Alzheimer's disease with high pathological and potential therapeutic relevance.

Nature (in press)

Willem M.; Tahirovic S.; Busche M.; Ovsepian S.; Chafai M.; Kootar S.; **Hornburg D.;** Evans L.; Moore S.; Daria A.; Hampel H.; Müller V.; Giudici C.; Nuscher B.; Wenninger-Weinzierl A.; Kremmer E.; Heneka M.; Thal D.; Giedraitis V.; Lannfelt L.; Müller U.; Livesey F.; Meissner F.; Herms J.; Konnerth A.; Marie H. and Haass C.

Contribution: This collaboration was established by the group of Christian Haass (Ludwig-Maximilian-University Munich). To characterize novel cleavage products of α/β/η-secretases I established protocols for partial proteolysis of these proteins using time courses and employed proteases of different specificities to map proteolytic events in APP processing. I prepared and measured the samples and analyzed the data. Furthermore, I wrote R scripts to calculate and visualize overlapping evidences from peptides generated in the experiments. In line with biochemical evidences, we were able to confirm and map novel APP processing mechanisms.

LETTER

η–Secretase processing of APP inhibits neuronal activity in the hippocampus

Michael Willem[1], Sabina Tahirovic[2], Marc Aurel Busche[3,4,5], Saak V. Ovsepian[2], Magda Chafai[6], Scherazad Kootar[6], Daniel Hornburg[7], Lewis D. B. Evans[8], Steven Moore[8], Anna Daria[1], Heike Hampel[1], Veronika Müller[1], Camilla Giudici[1], Brigitte Nuscher[1], Andrea Wenninger-Weinzierl[2], Elisabeth Kremmer[2,5,9], Michael T. Heneka[10,11], Dietmar R. Thal[12], Vilmantas Giedraitis[13], Lars Lannfelt[13], Ulrike Müller[14], Frederick J. Livesey[8], Felix Meissner[7], Jochen Herms[2], Arthur Konnerth[4,5], Hélène Marie[6] & Christian Haass[1,2,5]

Alzheimer disease (AD) is characterized by the accumulation of amyloid plaques, which are predominantly composed of amyloid-β peptide[1]. Two principal physiological pathways either prevent or promote amyloid-β generation from its precursor, β-amyloid precursor protein (APP), in a competitive manner[1]. Although APP processing has been studied in great detail, unknown proteolytic events seem to hinder stoichiometric analyses of APP metabolism *in vivo*[2]. Here we describe a new physiological APP processing pathway, which generates proteolytic fragments capable of inhibiting neuronal activity within the hippocampus. We identify higher molecular mass carboxy-terminal fragments (CTFs) of APP, termed CTF-η, in addition to the long-known CTF-α and CTF-β fragments generated by the α- and β-secretases ADAM10 (a disintegrin and metalloproteinase 10) and BACE1 (β-site APP cleaving enzyme 1), respectively. CTF-η generation is mediated in part by membrane-bound matrix metalloproteinases such as MT5-MMP, referred to as η-secretase activity. η-Secretase cleavage occurs primarily at amino acids 504–505 of APP$_{695}$, releasing a truncated ectodomain. After shedding of this ectodomain, CTF-η is further processed by ADAM10 and BACE1 to release long and short Aη peptides (termed Aη-α and Aη-β). CTFs produced by η-secretase are enriched in dystrophic neurites in an AD mouse model and in human AD brains. Genetic and pharmacological inhibition of BACE1 activity results in robust accumulation of CTF-η and Aη-α. In mice treated with a potent BACE1 inhibitor, hippocampal long-term potentiation was reduced. Notably, when recombinant or synthetic Aη-α was applied on hippocampal slices *ex vivo*, long-term potentiation was lowered. Furthermore, *in vivo* single-cell two-photon calcium imaging showed that hippocampal neuronal activity was attenuated by Aη-α. These findings not only demonstrate a major functionally relevant APP processing pathway, but may also indicate potential translational relevance for therapeutic strategies targeting APP processing.

To identify new proteolytic pathways of APP, we searched for CTFs other than those giving rise to P3 fragment (CTF-α) or amyloid-β (CTF-β)[3–5]. A new CTF with an approximate molecular mass of 30 kilodaltons (kDa) was revealed, which was recognized by an antibody to the C terminus of APP (Y188) and was absent in the brains of APP-knockout mice[6] (Fig. 1a and Supplementary Table 1). The molecular mass of the novel CTF suggests an additional physiological cleavage of APP amino-terminal to the known cleavage sites of β- and α-secretases, which we named accordingly η-cleavage of APP (Extended Data

Fig. 1). In the soluble fraction, we detected the N-terminal cleavage product (sAPP-η; Extended Data Fig. 2), with a molecular mass of approximately 80 kDa that distinguishes it from alternative N-terminal APP fragments described previously[7–9]. In addition, we observed lower molecular mass soluble peptides (Aη), which presumably derived from processing of CTF-η by BACE1 (Aη-β) or ADAM10 (Aη-α), or alternatively from sAPP-α/β cleavage (Fig. 1b). Aη was identified in the soluble fraction of mouse brains as several closely spaced peptides by antibody M3.2 (Fig. 1b), demonstrating that some of these fragments contain the N-terminal part of the amyloid-β domain and probably end at the α-secretase cleavage site. Aη fragments were further validated by antibody 9478D directed against an epitope N-terminal to the amyloid-β domain (Fig. 1b). Consistent with η-secretase cleavage of APP in wild-type mice, we observed increased CTF-η and Aη production in brain homogenates of APPPS1-21 transgenic mice[10] (Fig. 1c, d). Furthermore, antibody 192swe selectively identified the Aη-β species ending at the BACE1 cleavage site (Fig. 1d). Consistent with increased BACE1 cleavage of Swedish mutant APP, only minor amounts of Aη-α were detected in this mouse model with antibody 2D8 (Fig. 1d). Physiological η-secretase processing was further confirmed in cerebrospinal fluid (CSF) from humans with and without the Swedish mutation (APP$_{swe}$) (Fig. 1e). Fivefold higher Aη than amyloid-β levels were observed in human CSF (5.33 ± 1.39 times (mean ± s.d.) higher Aη than amyloid-β estimated by 2D8 blot signals; $n = 7$, $P > 0.0001$, Student's t-test). Using antibody 192swe, Aη-β was selectively detected in the CSF of patients with the Swedish mutation, whereas antibodies 2E9 and 2D8 detected Aη-α in all analysed samples (Fig. 1e). Moreover, while these peptides are generated by η-secretase cleavage N-terminal to the amyloid-β domain, they do not reach the γ-secretase site (see mass spectrometric analysis in Fig. 1f), demonstrating that they are different to the previously described N-terminally extended amyloid-β variants[11,12]. Membrane-bound matrix metalloproteinases such as MT1-MMP and MT5-MMP (also known as MMP14 and MMP24, respectively) were shown to cleave APP *in vitro* at a site consistent with the molecular mass of η-secretase processing products[13,14]. We therefore produced a neo-epitope-specific antibody (10A8; Extended Data Fig. 2) to identify the η-secretase cleavage site. Antibody 10A8 detected a protein corresponding to sAPP-η with an approximate molecular mass of 80 kDa in mouse brain lysates, which was absent in the APP-knockout mouse brain (Extended Data Fig. 2). Thus, η-secretase cleavage of APP may occur *in vivo* at least in part at amino

[1]Biomedical Center (BMC), Ludwig-Maximilians-University Munich, 81377 Munich, Germany. [2]German Center for Neurodegenerative Diseases (DZNE) Munich, 81377 Munich, Germany. [3]Department of Psychiatry and Psychotherapy, Technische Universität München, 81675 Munich, Germany. [4]Institute of Neuroscience, Technische Universität München, 80802 Munich, Germany. [5]Munich Cluster for Systems Neurology (SyNergy), Ludwig-Maximilians-University Munich, 81377 Munich, Germany. [6]Institut de Pharmacologie Moléculaire et Cellulaire (IPMC), Centre National de la Recherche Scientifique (CNRS), Université de Nice Sophia Antipolis, UMR 7275, 06560 Valbonne, France. [7]Max Planck Institute of Biochemistry, Martinsried 82152, Germany. [8]Gurdon Institute, Cambridge Stem Cell Institute & Department of Biochemistry, University of Cambridge, Cambridge CB2 1QN, UK. [9]Institute of Molecular Immunology, German Research Center for Environmental Health, 81377 Munich, Germany. [10]Department of Neurology, Clinical Neuroscience Unit, University of Bonn, 53127 Bonn, Germany. [11]German Center for Neurodegenerative Diseases (DZNE) Bonn, 53175 Bonn, Germany. [12]Institute of Pathology - Laboratory for Neuropathology, University of Ulm, 89081 Ulm, Germany. [13]Department of Public Health/Geriatrics, Uppsala University, 751 85 Uppsala, Sweden. [14]Institute for Pharmacy and Molecular Biotechnology IPMB, Functional Genomics, University of Heidelberg, 69120 Heidelberg, Germany.
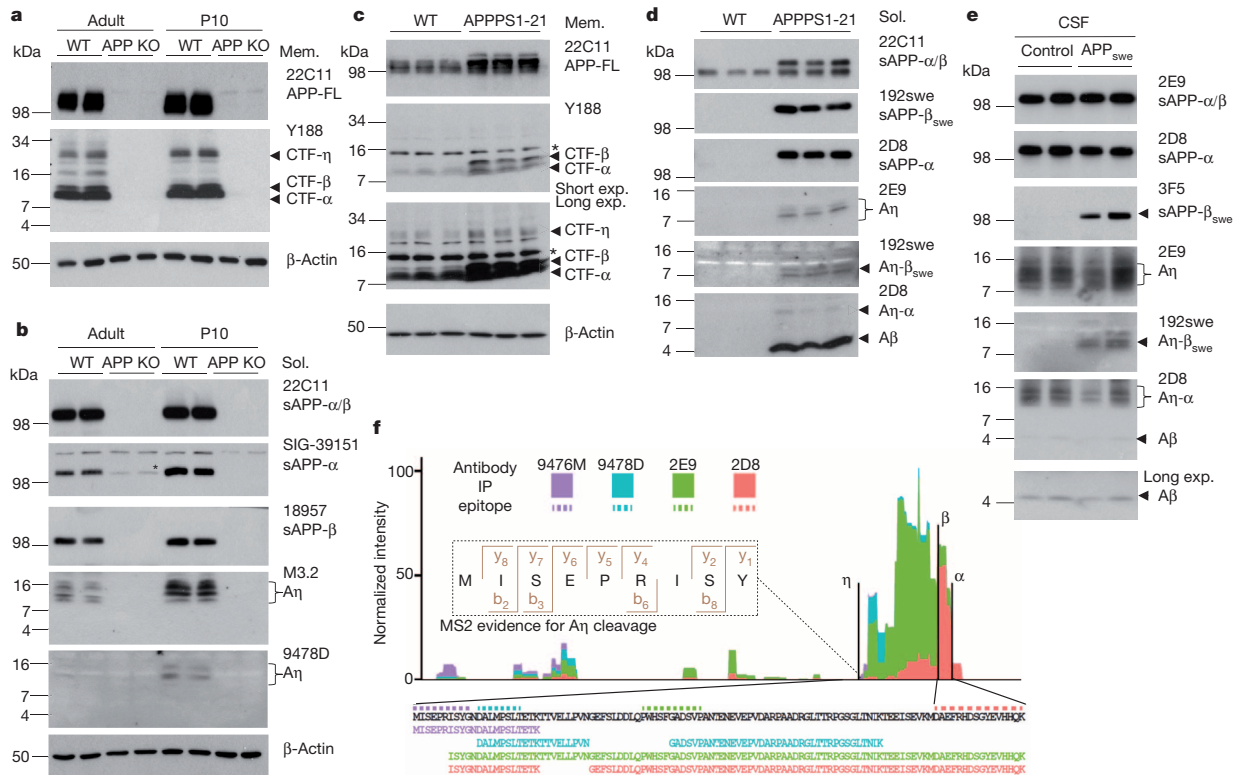
Figure 1 | A novel APP proteolytic processing pathway. a, A 30-kDa APP CTF-η is detected in the brains of adult and postnatal mice (P10, postnatal day 10). APP-FL, full-length APP; KO, knockout; WT, wild type. **b,** Aη was detected in the soluble (sol.) fraction of adult and P10 mice by antibody M3.2 and 9478D (antibody 9478D may not be sensitive enough to detect the lower Aη levels in adult brain). **c,** Higher levels of CTF-η are observed in APPPS1-21 mouse brains as compared to wild-type in the membrane (mem.) fraction. Short and long exposures indicated. Background bands are indicated by asterisks. **d,** Soluble extracts of APPPS1-21 mouse brains contained Aη species, as detected by 2E9 antibody. Aη-β$_{swe}$ was selectively detected by antibody 192swe. **e,** Aη and amyloid-β (Aβ) were readily detectable in human CSF by antibody 2D8. Antibody 2E9 allowed the detection of Aη in all samples, whereas 192swe specifically detected Aη-β$_{swe}$. **f,** Mass spectrometric analysis of Aη. Peptide intensities were summed per amino acid residue and plotted in relation to each other.

acids 504–505 of APP$_{695}$. To determine the N- and C-terminal cleavage sites of Aη peptides, we performed immunoprecipitation with antibodies 9476M, 9478D, 2D8 and 2E9 (Extended Data Fig. 3a, b). Isolated Aη peptides were digested with three different proteases to produce several overlapping peptides, and analysed by mass spectrometry. We identified several peptides covering the entire sequence between the cleavage site N504–M505 of APP$_{695}$ starting with the sequence MISEPRISY after the η-secretase cleavage site (Fig. 1f). Mass spectrometry also supports the C-terminal cleavages at the β- and α-secretase sites. After immunoprecipitation with 2D8, fragments of the amyloid-β domain were also observed in smaller amounts alongside the novel Aη peptides (Fig. 1f and Extended Data Fig. 3c). As the η-secretase cleavage site at amino acid 505 of APP is consistent with the previously described *in vitro* cleavage sites of APP$_{695}$ by MT1-MMP and MT5-MMP (refs 13, 14), we investigated brains from MT5-MMP- and MT1-MMP-knockout mice[15,16] (Extended Data Fig. 4). Whereas MT1-MMP knockout had no marked effect on Aη-α levels (Extended Data Fig. 4b), the generation of Aη-α was reduced in brains from MT5-MMP-knockout mice (Extended Data Fig. 4c). Furthermore, after MT5-MMP overexpression in murine N2a cells, a selective increase in Aη-α peptide of approximately 16 kDa was observed (data not shown). Thus, MT5-MMP displays η-secretase activity in intact mouse brains, although the contribution of other η-secretases must be considered.

While investigating protease inhibitors capable of blocking η-secretase, we observed that pharmacological BACE1 inhibition led to a pronounced accumulation of the long Aη-α species in Chinese hamster ovary (CHO) 7PA2 cells (Fig. 2a). This indicates that after blockade of β-secretase activity, processing by α-secretase leads to enhanced production of the long Aη-α species at the expense of the shorter BACE1-generated Aη-β. Similarly, BACE1 inhibition also led to an accumulation of endogenous CTF-η and enhanced production of endogenous Aη-α in primary mouse hippocampal neurons (Fig. 2b, c), as well as human neurons differentiated from embryonic pluripotent stem cells (H9 cells[17]; Fig. 2d–g). Furthermore, in human neurons we not only detected a 65% increase (Fig. 2g) in the slightly longer Aη-α species after BACE inhibition, but also a concomitant decrease in Aη-β peptides (Fig. 2e, top). Importantly, η-secretase processing significantly exceeded amyloidogenic processing (9.5 ± 1.87 times Aη compared to amyloid-β estimated for human neurons; $n = 8$, $P > 0.001$, Student's *t*-test). Pharmacological intervention with a BACE1 inhibitor *in vivo* caused a clear and time-dependent increase in CTF-η and Aη-α levels in APP$_{V717I}$ transgenic mice[18] (Fig. 2h), which was fully reversible within 24 h after administration. In agreement, increased CTF-η and Aη-α levels were also observed in a BACE1-knockout mouse[19] *in vivo* (Fig. 2i).

To identify a potential contribution of η-secretase processing to AD pathology, immunohistochemical analyses were performed with brains derived from six-month-old APPPS1-21 (ref. 20) mice (Extended Data Fig. 5a). This revealed co-labelling of antibody Y188 with antibody 2E9 in dystrophic neurites. No signal for Aη peptides was obtained in plaque cores in which aggregated amyloid-β was
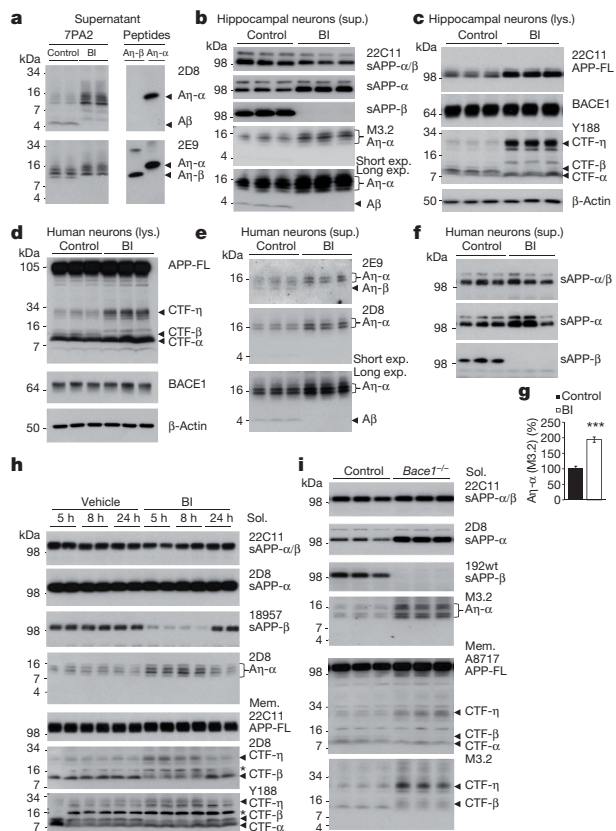
Figure 2 | Inhibition of BACE1 results in increased levels of CTF-η and Aη-α. **a**, Conditioned media from 7PA2 cells treated with or without a BACE1 inhibitor (BI) was co-migrated with synthetic Aη-β and Aη-α peptides and immunoblotted with 2D8 and 2E9 antibodies. **b–f**, After BACE inhibition in mouse hippocampal neurons and human neurons, a reduction of sAPP-β was accompanied by a strong increase in endogenous Aη-α and CTF-η levels. Lys., lysate; sup., supernatant. **g**, Quantification of intensities for 2D8 signals in **e** (*n* = 8; ****P* < 0.001, Student's *t*-test). **h**, BACE1 inhibition *in vivo* resulted in enhanced production of Aη-α species in APP_V717I mice. Background bands are indicated by asterisks. **i**, Western blot analysis of soluble extracts of P10 *Bace1*⁻ᐟ⁻ mouse brains revealed a marked increase in Aη-α peptides as compared to controls.

Figure 3 | Aη-α impairs hippocampal LTP. **a**, Pharmacological BACE1 inhibition with SCH1682496 lowers hippocampal LTP. DMSO, dimethylsulfoxide; fEPSP, field excitatory postsynaptic potential; HFS, high-frequency stimulation. **b**, Representative fEPSPs recorded in CA1 area before and 45 min after tetanization of Schaffer collaterals (top), with summary plot of the effects of the inhibitor (SCH1682496) and vehicle (DMSO) on fEPSP slopes in all examined groups in **a** (*n* = 9). **c**, Soluble Aη-α and Aη-β peptides were expressed in CHO cells and blotted with 2D8, 2E9 and 9478D antibodies. **d–f**, Aη-α (*n* = 9) (**d**), but not Aη-β (*n* = 7) (**e**), conditioned media from untransfected cells (CHO; *n* = 13) or artificial cerebrospinal fluid (ACSF) (control; *n* = 15) (**f**) significantly inhibited LTP. **g**, Summary graph of LTP magnitudes calculated 45–60 min after high-frequency stimulation from graphs in **d–f** with statistical analysis (**P* <0.05; one-way analysis of varaince (ANOVA) and post hoc Bonferroni test); error bars represent s.e.m. For each condition, sample fEPSP traces pre-LTP (black) and 45–60 min post-LTP (grey) induction are shown (top).

detected by 6E10 staining (Extended Data Fig. 5a). Similar data were obtained in human AD brains (Extended Data Fig. 6). To verify accumulation of CTF-η in dystrophic neurites, we used laser capture microdissection (LCM). Western blot analysis revealed not only CTF-β and CTF-α, but also CTF-η within the halo, but not within the plaque core area or regions devoid of plaques (Extended Data Fig. 5b). As expected, amyloid-β was observed within the plaque core as well as in the surrounding halo (Extended Data Fig. 5b).

Since the cleavage products of η-secretase APP processing accumulate after BACE1 inhibition and are enriched in dystrophic neurites, we examined whether soluble Aη peptides interfere with neuronal function, similar to soluble amyloid-β oligomers[1]. Long-term potentiation (LTP) is considered as a synaptic correlate of memory, and is widely used as a model for investigating the neurotoxic effects of amyloid-β oligomers on synaptic function[21–23]. A single oral dose of the BACE1 inhibitor SCH1682496 increased the CTF-η levels, and almost doubled the Aη-α level in soluble brain extracts prepared from animals 3 h after treatment (Extended Data Fig. 7). This was accompanied by a significant reduction of hippocampal LTP (Fig. 3a, b). These findings may suggest an involvement of Aη-α in LTP deficit

under acute blockade of β-secretase activity. To validate the potential effects of Aη peptides on synaptic transmission and plasticity directly, we expressed Aη-β and Aη-α in CHO cells (Fig. 3c). Concentrated conditioned media was further enriched for Aη by size-exclusion chromatography (SEC) and applied to hippocampal slices before LTP induction in CA1 pyramidal neurons. Neither Aη-β nor Aη-α influenced the baseline synaptic transmission (Extended Data Fig. 8). Comparison of LTP 60 min after its induction in the presence of Aη-β or Aη-α with control conditions (Fig. 3d–g) revealed that Aη-α lowered the LTP to a degree comparable to synthetic amyloid-β dimers[23] (Extended Data Fig. 9a, b), while truncated Aη-β had no effect (Fig. 3e, g). In support of this observation, synthetic Aη-α reduced LTP to a similar extent at a concentration of 100 nM (Extended Data Fig. 9c, d). To examine the direct effects of Aη peptides on neuronal activity *in vivo*, we used two-photon calcium imaging at single-cell resolution[24,25]. Figure 4a–d illustrates results from experiments in which the activity of neurons in the CA1 pyramidal cell layer was monitored before and after superfusion of the exposed hippocampus with Aη peptides or the respective control peptides. Aη-α strongly suppressed the activity of hippocampal neurons *in vivo*, an effect not observed with Aη-β or the control peptide (Fig. 4a–d and Extended Data Fig. 10). By using local application of synthetic Aη-α to hippocampal neurons, we demonstrate that the inhibitory effect of Aη-α on neurons was readily reversible after

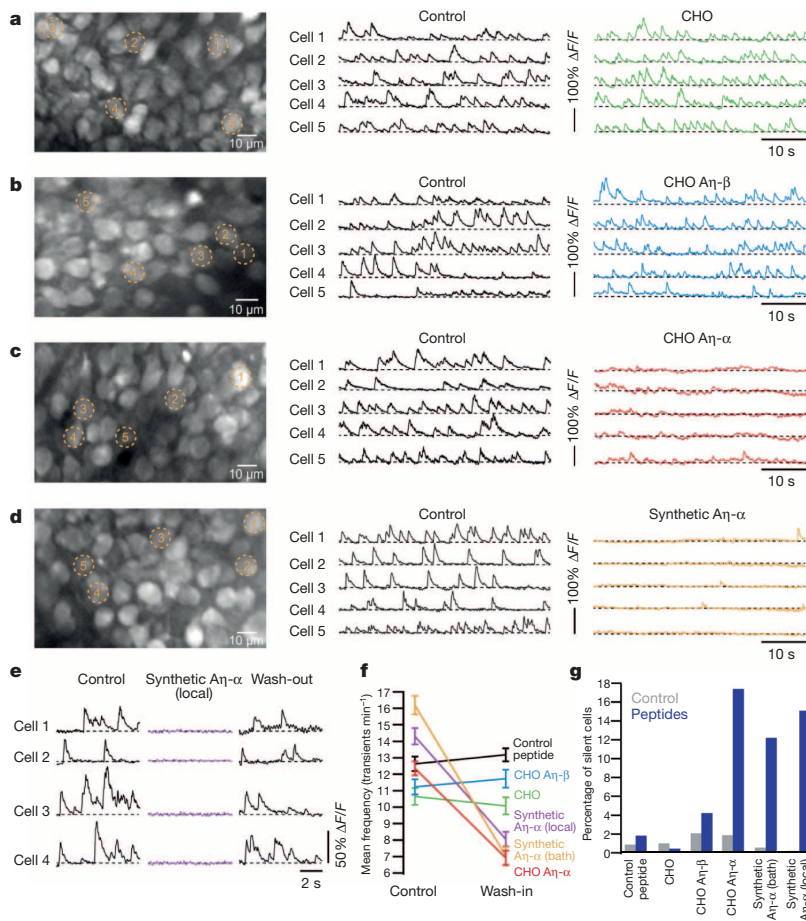**Figure 4 | Aη-α reduces neuronal activity *in vivo*. a–d**, Left, *in vivo* images of CA1 hippocampal neurons labelled with the fluorescent calcium indicator fluo-8AM. Middle and right, calcium transients of five representative neurons, marked in the corresponding left panels, before and during bath-application of Aη or CHO conditioned media. **e**, Calcium transients in hippocampal neurons before, during and after local application of synthetic Aη-α. **f, g**, Summary results of the changes in the average rates of calcium transients (**f**), and in the fractions of silent neurons (**g**); for statistics see Supplementary Tables 2 and 3. Error bars represent s.e.m.

washout (Fig. 4e). A summary of the results from all experiments is shown in Fig. 4f, g.

We have identified a new APP processing pathway that exceeds amyloidogenic processing. Similar to amyloid-β production[26], the alternative proteolytic processing pathway occurs under physiological conditions but may be altered during AD pathogenesis. Accumulation of η-secretase[27] and CTF-η within dystrophic neurites in close vicinity to neuritic plaques may also support its potential contribution to AD pathology. However, all APP and presenilin-associated familial AD mutations affect amyloid-β production and aggregation (reviewed in ref. 26), whereas the Icelandic mutation APP$_{A673T}$ prevents AD and dementia by moderately reducing amyloid-β production[28]. Indeed, the Swedish mutation decreased Aη-α by strongly enhancing BACE1-mediated APP processing. However, η-secretase accumulation[27] and its activity near amyloid plaques may suggest that η-secretase stimulation by amyloid-β could be a downstream effector within the amyloid cascade. Although Aη may be involved in the modulation of neuronal activity and synaptic plasticity, the differential bioactivity of recombinant Aη-α and Aη-β is currently unclear. One may speculate that the longer Aη-α peptide is more stable owing to unknown post-translational modifications. This would be consistent with our observation that, in contrast to cell-produced Aη-β, 100 nM of synthetic Aη-β inhibits LTP (data not shown). Finally, it is important to note that low-*n* amyloid-β oligomer preparations from 7PA2 supernatants contain considerable amounts of Aη (data not shown). Thus, the previously observed inhibition of LTP with such fractions may also be attributed to the presence of Aη. Our findings may also be considered in the context of continuing clinical trials with BACE1 inhibitors. Together with the identification of numerous brain-specific BACE1 substrates[29,30], our data indicate that therapeutic inhibition of BACE1 activity requires careful titration to prevent unwanted adverse effects at several levels.

1. Haass, C. & Selkoe, D. J. Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid β-peptide. *Nature Rev. Mol. Cell Biol.* **8,** 101–112 (2007).
2. Dobrowolska, J. A. *et al.* CNS amyloid-β, soluble APP-α and -β kinetics during BACE inhibition. *J. Neurosci.* **34,** 8336–8346 (2014).
3. Haass, C., Koo, E. H., Mellon, A., Hung, A. Y. & Selkoe, D. J. Targeting of cell-surface β-amyloid precursor protein to lysosomes: alternative processing into amyloid-bearing fragments. *Nature* **357,** 500–503 (1992).
4. Estus, S. *et al.* Potentially amyloidogenic, carboxyl-terminal derivatives of the amyloid protein precursor. *Science* **255,** 726–728 (1992).
5. Haass, C. *et al.* β-Amyloid peptide and a 3-kDa fragment are derived by distinct cellular mechanisms. *J. Biol. Chem.* **268,** 3021–3024 (1993).
6. Müller, U. *et al.* Behavioral and anatomical deficits in mice homozygous for a modified β-amyloid precursor protein gene. *Cell* **79,** 755–765 (1994).
7. Nikolaev, A., McLaughlin, T., O'Leary, D. D. & Tessier-Lavigne, M. APP binds DR6 to trigger axon pruning and neuron death via distinct caspases. *Nature* **457,** 981–989 (2009).
8. Jefferson, T. *et al.* Metalloprotease meprin β generates nontoxic N-terminal amyloid precursor protein fragments *in vivo. J. Biol. Chem.* **286,** 27741–27750 (2011).
9. Vella, L. J. & Cappai, R. Identification of a novel amyloid precursor protein processing pathway that generates secreted N-terminal fragments. *FASEB J.* **26,** 2930–2940 (2012).
10. Radde, R. *et al.* Aβ42-driven cerebral amyloidosis in transgenic mice reveals early and robust pathology. *EMBO Rep.* **7,** 940–946 (2006).

11. Portelius, E. *et al.* Mass spectrometric characterization of amyloid-β species in the 7PA2 cell model of Alzheimer's disease. *J. Alzheimers Dis.* **33**, 85–93 (2013).
12. Welzel, A. T. *et al.* Secreted amyloid β-proteins in a cell culture model include N-terminally extended peptides that impair synaptic plasticity. *Biochemistry* **53**, 3908–3921 (2014).
13. Higashi, S. & Miyazaki, K. Novel processing of β-amyloid precursor protein catalyzed by membrane type 1 matrix metalloproteinase releases a fragment lacking the inhibitor domain against gelatinase A. *Biochemistry* **42**, 6514–6526 (2003).
14. Ahmad, M. *et al.* Cleavage of amyloid-β precursor protein (APP) by membrane-type matrix metalloproteinases. *J. Biochem.* **139**, 517–526 (2006).
15. Folgueras, A. R. *et al.* Metalloproteinase MT5-MMP is an essential modulator of neuro-immune interactions in thermal pain stimulation. *Proc. Natl Acad. Sci. USA* **106**, 16451–16456 (2009).
16. Zhou, Z. *et al.* Impaired endochondral ossification and angiogenesis in mice deficient in membrane-type matrix metalloproteinase I. *Proc. Natl Acad. Sci. USA* **97**, 4052–4057 (2000).
17. Shi, Y., Kirwan, P. & Livesey, F. J. Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks. *Nature Protocols* **7**, 1836–1846 (2012).
18. Moechars, D. *et al.* Early phenotypic changes in transgenic mice that overexpress different mutants of amyloid precursor protein in brain. *J. Biol. Chem.* **274**, 6483–6492 (1999).
19. Cai, H. *et al.* BACE1 is the major β-secretase for generation of Aβ peptides by neurons. *Nature Neurosci.* **4**, 233–234 (2001).
20. Radde, R. *et al.* Aβ42-driven cerebral amyloidosis in transgenic mice reveals early and robust pathology. *EMBO Rep.* **7**, 940–946 (2006).
21. Walsh, D. M. *et al.* Naturally secreted oligomers of amyloid β protein potently inhibit hippocampal long-term potentiation *in vivo. Nature* **416**, 535–539 (2002).
22. Shankar, G. M. *et al.* Natural oligomers of the Alzheimer amyloid-β protein induce reversible synapse loss by modulating an NMDA-type glutamate receptor-dependent signaling pathway. *J. Neurosci.* **27**, 2866–2875 (2007).
23. Shankar, G. M. *et al.* Amyloid-β protein dimers isolated directly from Alzheimer's brains impair synaptic plasticity and memory. *Nature Med.* **14**, 837–842 (2008).
24. Busche, M. A. *et al.* Clusters of hyperactive neurons near amyloid plaques in a mouse model of Alzheimer's disease. *Science* **321**, 1686–1689 (2008).
25. Busche, M. A. *et al.* Critical role of soluble amyloid-β for early hippocampal hyperactivity in a mouse model of Alzheimer's disease. *Proc. Natl Acad. Sci. USA* **109**, 8740–8745 (2012).
26. Haass, C. Take five–BACE and the γ-secretase quartet conduct Alzheimer's amyloid β-peptide generation. *EMBO J.* **23**, 483–488 (2004).
27. Sekine-Aizawa, Y. *et al.* Matrix metalloproteinase (MMP) system in brain: identification and characterization of brain-specific MMP highly expressed in cerebellum. *Eur. J. Neurosci.* **13**, 935–948 (2001).
28. Jonsson, T. *et al.* A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* **488**, 96–99 (2012).
29. Kuhn, P. H. *et al.* Secretome protein enrichment identifies physiological BACE1 protease substrates in neurons. *EMBO J.* **31**, 3157–3168 (2012).
30. Zhou, L. *et al.* The neural cell adhesion molecules L1 and CHL1 are cleaved by BACE1 protease *in vivo. J. Biol. Chem.* **287**, 25927–25940 (2012).

METHODS

Cell culture. Mycoplasm-free CHO and 7PA2 (ref. 31) cells (gift from E. Koo) were grown in DMEM/F12 (Thermo Scientific) supplemented with 10% FCS (Thermo Scientific) plus penicillin/streptomycin and non-essential amino acids (PAA) in a humid incubator with 5% CO_2 at a temperature of 37 °C. For inhibitor treatment, cell culture medium was replaced with fresh, pre-warmed serum free medium (OPTIMEM; Thermo Scientific) supplemented with inhibitors or DMSO as vehicle control. Treatment was initiated when cells reached 90–100% confluency and conditioned media were collected after 20–24 h. Supernatants were cleared by centrifugation (10 min, 5,500g at 4 °C). To obtain cell lysates, cell monolayers were washed once with ice-cold PBS and detached in 1 ml PBS using a cell scraper. The cell suspension was pelleted by centrifugation (5 min, 1,000g at 4 °C) and lysed with RIPA buffer (20 mM sodium citrate, pH 6.4, 1 mM EDTA, 1% Triton X-100 in ddH_2O) supplemented with Protease Inhibitor Cocktail (Sigma-Aldrich). The protein concentration of lysates was determined using the Uptima BC Assay Protein Quantitation kit (Interchim).

Primary cell culture. Hippocampal neurons were isolated from embryonic day 18 CD rats (Charles River) as described previously[32]. Dissociated neurons were plated at 17,700 cells cm^{-2} onto 6-cm dishes coated with poly-L-lysine (1 mg ml^{-1}; Sigma) and cultured in Neurobasal medium supplemented with 2% B27 and 0.5 mM L-glutamine (all from Invitrogen). Hippocampal cultures were maintained in a humidified 5% CO_2 incubator at 37 °C. For inhibitor treatment, DIV16 culture medium was replaced with fresh, pre-equilibrated N2 medium (supplemented with 20% of 4 days conditioned N2 medium from pure primary cultured astrocytes) to which inhibitors or DMSO as vehicle control were added.

Generation and BACE1 inhibition of cerebral cortex neurons induced from human embryonic stem cells. Cell lines in this study were H9 ES (WiCell Research Institute)[33]. Pluripotent cells were cultured on mouse embryonic fibroblasts (GlobalStem) in DMEM/F12 containing 20% (v/v) KSR, 100 μm non-essential amino acids, 100 μM 2-mercaptoethanol, 50 U ml^{-1} penicillin and 50 mg ml^{-1} streptomycin (Life Technologies) and 10 mg ml^{-1} FGF2. Directed differentiation of human embryonic stem cells to cerebral cortex neurons was carried out as described[17,34]. Human neurons (75 days after induction) were treated with 1 μM β-secretase inhibitor LY2886721 (Selleck) dissolved in DMSO (20 mM stock). Vehicle-only control assays were performed using DMSO. The compound was applied twice at 48-h intervals. Extracellular media was collected before drug addition and at subsequent 48-h intervals. Neurons were collected after 4 days of treatment using 0.5 mM EDTA in PBS.

Transgenic mice, animal care and animal handling. $Bace1^{-/-}$ and APPPS1-21 mice were described before[10,19] and were bred for this study in a Bl6C57/J background. All treatments were approved by the local committee for animal use and were performed in accordance to state and federal regulations (license number KVR-I/221-TA116/09). Mice had access to pre-filtered sterile water and standard mouse chow (Ssniff Ms-H, Ssniff Spezialdiäten GmbH, Soest, Germany) *ad libitum* and were housed under a reversed day–night rhythm in IVC System Type II L-cages (528 cm^2) equipped with solid floors and a layer of bedding, in accordance to local legislation on animal welfare.

BACE1 inhibitor treatment. Randomized APP_{V717I} (ref. 18) mice were treated with vehicle or with the inhibitor RO5508887 provided by Hoffmann-La Roche[35]. The groups of treated mice were blinded to the examiner and uncoded at the end of the experiments. Three-month-old heterozygous female transgenic mice in mixed FVB/N × C57Bl/6J background expressing human APP_{V717I} (ref. 18) were used for BACE1 inhibition studies. Gavage mediated administration of BACE1 inhibitor (90 mg kg^{-1}, 14.06 ml kg^{-1}) or vehicle (14.06 ml kg^{-1}) was performed once[35]. The BACE1 inhibitor was diluted in 5% ethanol (Merck) and 10% solutol (Sigma-Aldrich) in sterile water (Baxter). Animals were sacrificed after 5, 8 and 24 h. Mice were anaesthetized with 3.5 μl per gram body weight of a mixture of ketamine (115 mg ml^{-1} ketamine hydrochloride, Eurovet), xylazin 2% (23.32 mg ml^{-1} xylazine hydrochloride, VMD Arendonk), atropine (0.50 mg ml^{-1} atropine sulphate, Sterop) and saline (8:5:2:5, v/v/v/v). For brain preparation, mice were flushed *trans*-cardially with ice-cold saline (3.5 ml min^{-1}, 3 min). The brain was removed from the cranium and dissected into left and right hemiforebrain, brainstem, cerebellum and olfactory bulb. The brain structures were promptly immersed in liquid nitrogen and stored at −80 °C. Different tissues (kidneys, spleen, liver, stomach, gut, lungs and heart) were examined and checked for gross abnormalities. No obvious abnormalities were observed in any of the treatment groups.

Preparation of protein extracts from brain. Brains were removed from the cranium and dissected into left and right hemispheres. Brain tissue was snap-frozen in liquid nitrogen and stored at −80 °C. Soluble proteins were extracted with DEA buffer (50 mM NaCl, 0,2% diethylamine, pH 10, plus protease inhibitor (P8340, Sigma-Aldrich))[36], membrane proteins were extracted with RIPA buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM EGTA,1% NP-40,1% sodium deoxycholate, 2.5 mM sodium pyrophosphate

plus protease inhibitor) or applying a membrane preparation protocol as described before[37].

Protein analysis. Proteins were separated under denaturing conditions using discontinuous SDS–PAGE. Equal amounts of proteins denatured in Laemmli buffer were loaded onto the gel and 10 μl of the SeeBlue Plus2 Prestained Standard (Invitrogen) served as molecular mass marker. Electrophoresis was performed in Tris-glycine buffer (25 mM Tris, 190 mM glycine in ddH_2O) using the Mini-PROTEAN system (BIORAD) on activated PVDF membranes. Low molecular mass proteins (<16 kDa) were separated using precast gradient Tricine Protein Gels (10–20%, 1 mm, Novex) in Tris-tricine buffer using the XCell SureLock Mini-Cell system (Novex). After separation by SDS–PAGE, proteins were transferred onto membranes using the tank/wet Mini Trans-Blot cell system (BIORAD). CTFs, Aη and amyloid-β were detected after transfer on Nitrocellulose membranes (Protran BA85; GE Healthcare), while other proteins were blotted on PVDF (Immobilon-P, Merck Millipore). As size markers for Aη synthetic peptides Aη-β (92 amino acids; 1-MISEPRISYGNDALMPSLTETKT TVELLPVNGEFSLDDLQPWHSFGADSVPANTENEVEPVDARPAADRGLTT RPGSGLTNIKTEEISEVKM-92) and the slightly longer Aη-α (108 amino acids; 1-MISEPRISYGNDALMPSLTETKTTVELLPVNGEFSLDDLQPWHSFGADSV PANTENEVEPVDARPAADRGLTTRPGSGLTNIKTEEISEVKMDAEFRHDSG YEVHHQK-108) were obtained from Peptide Speciality Laboratories. After completion of the transfer and before blocking, proteins transferred to nitrocellulose membranes were additionally denatured by boiling the membrane in PBS (140 mM NaCl, 10 mM Na_2HPO_4, 1.75 mM KH_2PO_4, 2.7 mM KCl in ddH2O, pH 7.4) for 5 min. After cooling to room temperature, nitrocellulose membranes as well as the PVDF membranes were blocked in I-Block solution (0.2% Tropix I-Block (Applied Biosystems), 0.1% Tween20 in PBS) for 1 h at room temperature or overnight at 4 °C (with agitation). Transferred proteins were detected using immunodetection and enhanced chemiluminescence (ECL). First, blocked membranes were incubated with primary antibodies diluted in I-Block solution overnight at 4 °C (with agitation). After removal of the antibody, membranes were washed three times in TBS-T buffer (10 min each, at room temperature, with agitation; 140 mM NaCl, 2.68 mM KCl, 24.76 mM Tris, 0.3% Triton X-100 in ddH_2O, pH 7.6) and subsequently incubated with a horseradish-peroxidase-coupled secondary antibody (obtained from Promega or Santa Cruz). Secondary antibodies were diluted in I-Block solution and membranes were incubated for 1 h at room temperature (with agitation) followed by three washes in TBS-T. For ECL detection, membranes were incubated with horseradish peroxidase substrate (ECL, GE Healthcare or ECL Plus, Thermo Scientific) for 1 min at room temperature and signals were captured with X-ray films (Super RX Medical X-Ray, Fujifilm), which were subsequently developed using an automated film developer (CAWOMAT 2000 IR, CAWO). Quantitation of protein was conducted using ImageJ software. Ratios were obtained from signals on the same film for Aη over amyloid-β. Quantitative data were analysed statistically by using a two-tailed Student's t-test.

Molecular cloning and transfection. For the expression of Aη-α and Aη-β in CHO cells, the complementary DNAs of the respective fragments were amplified by PCR and subcloned into the pSecTag2A (Invitrogen) vector that features an N-terminal secretion signal. CHO cells were cultured in DMEM with 10% FCS and non-essential amino acids. Transfections were carried out using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions.

Mass spectrometry analysis of samples. Beads with immunoprecipitated peptides were resuspended in ddH_2O and reduced with 10 mM dithiothreitol followed by alkylation with 55 mM 2-chloroacetamide. Samples were divided into three parts and digested with either 1 μg of trypsin (1.2 M urea, 0.4 M thiourea and 50 mM ammonium bicarbonate), LysC (5 M urea, 1.7 M thiourea and 50 mM ammonium bicarbonate) or chymotrypsin (0.3 M urea, 0.1 M thiourea and 50 mM ammonium bicarbonate). To increase the sequence coverage further, partially cleaved peptides were generated by digesting for 5, 10, 20, 40, 60, 120, 180 and 720 min. Samples from all time points of a respective protease were pooled and desalted on stage tips[38].

For liquid chromatography–tandem mass spectrometry (LC–MS/MS), peptides were separated on a Thermo Scientific EASY-nLC 1000 HPLC system (Thermo Fisher Scientific) and in-house packed columns (75 μm inner diameter, 20 cm length, 1.9 μm C18 particles (Dr. Maisch GmbH). The peptide mixture was loaded in buffer A (0.5% formic acid) and separated with a gradient from 10% to 60% buffer B (80% acetonitrile, 0.5% formic acid) within 40 min at 250 nl min^{-1} at a column temperature of 50 °C. A Quadrupole Orbitrap mass spectrometer[39] (Q Exactive, Thermo Fisher Scientific) was coupled to the HPLC system via a nano electrospray source. We used data-dependent acquisition with a survey scan range of 300 to 1,650 m/z, at a resolution of 60,000 m/z and selected up to five most abundant features with a charge state ≥ 2 for HCD fragmentation[40] at a normalized collision energy of 27 and a resolution of 15,000 at m/z 200. To limit repeated

sequencing, dynamic exclusion of sequenced peptides was set to 20 s. Thresholds for ion injection time and ion target values were set to 20 ms and 3×10^6 for the survey scans, and 120 ms and 1×10^5 for the MS/MS scans. Data were acquired using the Xcalibur software (Thermo Scientific).

Data analysis. To process mass spectrometry raw files, we used the MaxQuant software (v1.5.2.16)[41]. We used the Andromeda search engine[42], which is integrated into MaxQuant, to search MS/MS spectra against the APP$_{695}$ and 247 common contaminating proteins[42]. We set enzyme specificity to unspecific to detect novel cleavage sites and set a peptide search length from 7 to 40 amino acids. A false discovery rate cutoff of 1% was applied at the peptide level. For data visualization we used R[43]. Identified peptides were mapped to APP$_{695}$. To display quantitative evidence for overlapping peptides, intensities of identified peptides were summed and plotted per amino acid residue. The data of the individual immunoprecipitation and mass spectrometry analyses are depicted in Extended Data Fig. 3.

Human CSF samples. Human CSF samples collected at the Department of Neurology Outpatient unit for neurodegenerative disease (KBFZ) of the University of Bonn were obtained by lumbar puncture at position L3, centrifuged and divided in small aliquots. For further analysis, samples were stored at $-80\,^\circ$C. Turbid or blood-contaminated samples were excluded from analysis. Use of these samples for research purposes has been consented by all patients according to the ethical committee requirements of the University of Bonn Ethical committee and approval number 279/10. For the analysis of APP$_{swe}$ carriers with antibodies 192swe (ref. 44) lumbar CSF was obtained from family members. Tubes with CSF were stored at $-70\,^\circ$C until analysis. The clinical diagnosis of probable AD was based on NINCDS-ADRDA criteria[45]. The diagnosis of AD was confirmed by neuropathological examination of the brain of one deceased mutation-carrier[46,47]. This study was approved by research ethics committee at the Uppsala University Hospital (Dnr 048-2005).

Neuropathology and immunohistochemistry. Use of brain samples for research purposes has been consented by all patients according to the ethical committee requirements of the University of Ulm Ethical committee and approval number 54/08. Braak-NFT stages[48], and CERAD[49] scores for neuritic plaques were used to determine the degree of AD pathology according to the NIA-AA guidelines[50]. Consecutive paraffin sections from the human medial lobe were stained with 22C11, 9476M and 9478D. Primary antibodies were detected with biotinylated anti-mouse and anti-rabbit IgG secondary antibodies and visualized with avidin-biotin-complex (ABC-Kit, Vector Laboratories) and diaminobenzidine-HCl (DAB). The sections were counterstained with haematoxylin. Positive and negative controls were performed. 9476M and 9478D stainings were assessed in 10 control and 10 AD patient cases.

For double immunofluorescence analysis of APPPS1-21 brain sections, 6-month-old mice were killed by CO_2 inhalation according to animal handling laws. Brains were dissected and fixed with 4% paraformaldehyde in 0.1 M PBS, pH 7.4 for 48 h. For immunohistochemistry, 25-μm-thick sagittal mouse brain cryosections were treated with 10 mM sodium citrate, pH 6 at 95 $^\circ$C for 20 min, washed with 0.5% Triton X-100 in PBS, blocked with 5% goat serum (Invitrogen) and 0.5% Triton X-100 in PBS for 1 h and subsequently incubated overnight with primary antibodies diluted in blocking solution. Primary antibodies were used as listed in Supplementary Table 1. DAPI was used to counterstain nuclei. Signals were visualized using fluorescently labelled secondary antibodies. Confocal images were acquired using a Plan-Apochromat 25×/0.8 oil differential interference contrast objective on a LSM 710 confocal microscope (Zeiss) in sequential scanning mode using ZEN 2011 software package (black edition, Zeiss).

LCM of plaque enriched brain material. For laser capture microdissection of plaque cores and halos, 10-, 11-, 14-, 16- and 24-month-old transgenic APPPS1-21 mice were used according to a previously published protocol[51] with slight modifications. Mice brains were dissected and immediately frozen on crushed dry ice. Ten-micrometre-thick sagittal sections were cut using a Microm HM 560 cryostat (Thermo Scientific), mounted on frame slides containing a 1.4 μm polyethylene terephthalate membrane (Leica Microsystems) and subsequently stained or stored at $-80\,^\circ$C for later usage. Staining was performed as follows: brain sections were thawed briefly at room temperature, fixed with 75% ethanol for 1 min, stained with 0.05% Thioflavin-S for 5 min, washed with 75% ethanol and dried at room temperature. LCM was performed on the same day using a laser dissection microscope (Leica, LMD 7000) with the following settings: excitation wavelength 495 nm, laser power 30, aperture 5, speed 6 and pulse frequency 119. From each animal, at least 800 plaque cores and halos, dissected from 12 brain sections were cut using a 63× magnification objective, collected in 0.5 ml caps (Leica Microsystem) and subsequently pooled for protein analysis. Areas containing no plaques were cut using a 10× magnification objective and were used as controls. Protein lysates were done essentially as described above using RIPA with 0.1% SDS.

Slice preparation and electrophysiological recordings applying Aη peptides *in vitro*. Transverse hippocampal slices (350 μm) were prepared from P20–30 Swiss mice following standard procedures[52]. Slices were cut in ice-cold oxygenated (95% O_2, 5% CO_2) solution containing 206 mM sucrose, 2.8 mM KCl, 1.25 mM NaH$_2$PO$_4$, 2 mM MgSO$_4$, 1 mM MgCl$_2$, 1 mM CaCl$_2$, 26 mM NaHCO$_3$, 0.4 mM sodium ascorbate and 10 mM glucose, pH 7.4. For recovery (1 h), slices were incubated at 27 $^\circ$C in oxygenated standard ACSF containing: 124 mM NaCl, 2.8 mM KCl, 1.25 mM NaH$_2$PO$_4$, 2 mM MgSO$_4$, 3.6 mM CaCl$_2$, 26 mM NaHCO$_3$, 0.4 mM sodium ascorbate, 10 mM glucose (pH 7.4)[53]. Slices were inspected in a chamber on an upright microscope (Slicescope, Scientifica Ltd) with infrared differential interference contrast illumination, and were perfused with the oxygenated ACSF at 27 ± 1 $^\circ$C. fEPSPs were recorded in the stratum radiatum of the CA1 region using a glass electrode (filled with 1 M NaCl, 10 mM HEPES, pH 7.4) and the stimuli (30% of maximal fEPSP) were delivered to the Schaeffer Collateral pathway by a monopolar glass electrode (filled with ACSF). Electrodes were specifically placed just below the surface of the slice to maximize the exposure to circulating peptides. A minimum of 15–20 min stable baseline was first obtained in standard ACSF followed by another 15–20 min of bath application of ACSF containing SEC fractions (CHO, Aη-α or Aη-β; 1/15 dilution, inter-leaved recordings) using re-circulation with a peristaltic pump at 2.5–3 ml min^{-1} while being continuously aerated with 95% oxygen. No alterations in fEPSP baseline responses were observed after incubation with the SEC fractions (Extended Data Fig. 8). In the continuous presence of ACSF/SEC solution, LTP was induced using a high-frequency stimulation protocol with two pulses of 100 Hz for 1 s with a 20 s interval between pulses, and recorded for 1 h. Control recordings (no application of SEC fractions) were obtained in an interleaved fashion in which ACSF was re-circulated using an identical procedure. For LTP analysis, the first third of the fEPSP slope was calculated in baseline condition (15–20 min before induction of LTP) and compared to that after LTP induction (60 min after tetanization of Schaffer collaterals). The average baseline value was normalized to 100% and all values of the experiment were normalized to this baseline average (1-min bins). Experimental data were pooled per condition and presented as mean ± s.e.m. Data analysis was performed with the Clampfit software (Molecular Devices). The test samples were blinded to the investigator and uncoded at the end of the experiments. Statistical analysis was performed using GraphPad (Prism 6) with the last 15 min of the recordings compared to measurements of 15–20 min of baseline, using a two-tailed Student's *t*-test for statistical analysis on two samples or one-way ANOVA and post hoc Bonferroni test for statistical analysis on three and more samples, with $P < 0.05$ taken as statistically significant. No power analysis was done to estimate sample size, and there was no randomization.

Electrophysiological recordings of the effects of BACE1 inhibitor *in vitro*. BACE1 inhibitor (100 mg kg^{-1}, single gavage) or vehicle-treated mice (12 weeks old) were deeply anaesthetized with isoflurane (1% in O_2) and decapitated with brains rapidly extracted and placed for 5–6 min in ice-cold bubbled (95% O_2, 5% CO_2) slicing solution (in mM): 75 sucrose, 85 NaCl, 2.5 KCl, 1.25 NaH$_2$PO$_4$, 25 NaHCO$_3$, 0.5 CaCl$_2$, 4 MgCl$_2$, 25 glucose, pH 7.4. Coronal slices (400 μm) containing the hippocampus were cut (VT1200S; Leica) and transferred into a warming chamber (35 $^\circ$C) filled with bubbled solution of the same composition, except sucrose was omitted and NaCl increased to 125 mM (30 min). This was followed by the transfer of slices into recording ACSF (in mM): 125 NaCl, 2.5 KCl, 1.25 NaH$_2$PO$_4$, 25 NaHCO$_3$, 2 CaCl$_2$, 2 MgCl$_2$, 25 glucose. Recordings of fEPSP were made from the hippocampal CA1 area. A glass bipolar stimulating electrode was placed in the stratum radiatum of CA2–CA3 subfields to stimulate Schaffer collaterals with 0.2 ms current pulses at 0.033 Hz (A-360, WPI), with evoked responses recorded in the stratum radiatum of CA1 area. Incrementing current pulses (0.2 mA) were used for obtaining stimulus-response relationship graphs. Stable baseline and LTP recordings were made using one-half of the maximal stimulus intensities; LTP was induced by high-frequency stimulation of Schaffer collaterals with 10 trains of 10 pulses at 100 Hz applied, with 2-s inter-train intervals. Signals were filtered at 5 kHz, digitally sampled at 10 kHz and stored for offline analysis. The relative slope and peak amplitude of evoked fEPSPs were measured using FitMaster (HEKA Electronics). The groups of treated mice were blinded to the investigator and uncoded at the end of the experiments. A one-way ANOVA and post hoc Bonferroni test have been used for statistical analysis, with $P < 0.05$ taken as statistically significant.

***In vivo* two-photon Ca^{2+} imaging.** All experimental procedures were in compliance with institutional animal welfare guidelines and were approved by the state government of Bavaria, Germany. The animal preparation procedure was similar to that described previously[25]. In brief, C57Bl/6 mice (male or female, ~P40) were anaesthetized with isoflurane (1–1.5%) and placed onto a warming plate (37–38 $^\circ$C). The skin was removed and a custom-made recording chamber was glued to the exposed skull. A craniotomy (~1 mm) was made over the

hippocampus (2.5 mm posterior to bregma, 2.2 mm lateral to the midline) and a small portion of the overlying cortical tissue was carefully removed by aspiration. The animal was placed under a microscope on a warm heating plate (37–38 °C) and kept anaesthetized with low-levels of isoflurane (~0.8%). Respiratory and pulse rates were continuously monitored. The recording chamber was perfused with warm normal Ringer's solution containing 125 mM NaCl, 4.5 mM KCl, 26 mM NaHCO$_3$, 1.25 mM NaH$_2$PO$_4$, 2 mM CaCl$_2$, 1 mM MgCl$_2$ and 20 mM glucose (pH 7.4 when bubbled with 95% O$_2$ and 5% CO$_2$). The exposed CA1 region of the hippocampus was then stained with fluo-8AM (AAT Bioquest; 0.6 mM) using the multi-cell bolus loading technique[54].

In vivo imaging was performed with a custom-built two-photon microscope equipped with a Ti:sapphire laser system (Coherent; laser wavelength 925 nm), a resonant scanner and a Pockel's cell for laser intensity modulation. Full-frame images were acquired at 30 Hz using a water-immersion objective (Nikon; 40×, 0.8 numerical aperture). Data acquisition was controlled using custom-written software based on LabVIEW (National Instruments). Image analysis was performed off-line by using custom routines in LabVIEW and Igor Pro (Wavemetrics). Cellular regions of interest were drawn around individual somata, and then relative fluorescence change ($\Delta F/F$) versus time traces were generated for each region of interest. Ca^{2+} transients were identified as changes in $\Delta F/F$ that were three times larger than the s.d. of the noise band.

To assess the effects of Aη peptides on neuronal activity *in vivo*, the peptides or the respective controls (SEC fractions obtained from untransfected CHO cells or a synthetic peptide (46 amino acids; 1-ADSVPANTENEVEPVDARPAADRGL TTRPGSGLTNIKTEEISEVKM-46) of a middle part of Aη were added to the normal Ringer's solution used for perfusion of the recording chamber (bath-application technique; 45–60 min each wash-in). In a subset of experiments, synthetic Aη-α (92 amino acids; 1-MISEPRISYGNDALMPSLTETKTTVELLPV NGEFSLDDLQPWHSFGADSVPANTENEVEPVDARPAADRGLTTRPGSGLT NIKTEEISEVKM-92) was applied locally by gentle pressure injection through a glass pipette that was placed close to the neurons of interest (local application technique; 40 s each pressure injection).

The samples were blinded to the investigator and uncoded at the end of the experiments. Statistical analysis was performed using SPSS. The statistical methods used were the Student's *t*-test and the Fisher's exact test. $P < 0.05$ was considered statistically significant.

31. Podlisny, M. B. *et al.* Aggregation of secreted amyloid β-protein into sodium dodecyl sulfate-stable oligomers in cell-culture. *J. Biol. Chem.* **270,** 9564–9570 (1995).
32. Kaech, S. & Banker, G. Culturing hippocampal neurons. *Nature Protocols* **1,** 2406–2415 (2006).
33. Israel, M. A. *et al.* Probing sporadic and familial Alzheimer's disease using induced pluripotent stem cells. *Nature* **482,** 216–220 (2012).
34. Shi, Y., Kirwan, P., Smith, J., Robinson, H. P. & Livesey, F. J. Human cerebral cortex development from pluripotent stem cells to functional excitatory synapses. *Nature Neurosci.* **15,** 477–486 (2012).
35. Jacobsen, H. *et al.* Combined treatment with a BACE inhibitor and anti-Ab antibody gantenerumab enhances amyloid reduction in APPLondon mice. *J. Neurosci.* **34,** 11621–11630 (2014).
36. Nolan, R. L. & Teller, J. K. Diethylamine extraction of proteins and peptides isolated with a mono-phasic solution of phenol and guanidine isothiocyanate. *J. Biochem. Biophys. Methods* **68,** 127–131 (2006).
37. Westmeyer, G. G. *et al.* Dimerization of β-site β-amyloid precursor protein-cleaving enzyme. *J. Biol. Chem.* **279,** 53205–53212 (2004).
38. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature Protocols* **2,** 1896–1906 (2007).
39. Scheltema, R. A. *et al.* The Q Exactive HF, a Benchtop mass spectrometer with a pre-filter, high-performance quadrupole and an ultra-high-field Orbitrap analyzer. *Mol. Cell. Proteomics* **13,** 3698–3708 (2014).
40. Olsen, J. V. *et al.* Higher-energy C-trap dissociation for peptide modification analysis. *Nature Methods* **4,** 709–712 (2007).
41. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnol.* **26,** 1367–1372 (2008).
42. Cox, J. *et al.* Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10,** 1794–1805 (2011).
43. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, http://www.R-project.org/ (2014).
44. Haass, C. *et al.* The Swedish mutation causes early-onset Alzheimer's-disease by β-secretase cleavage within the secretory pathway. *Nature Med.* **1,** 1291–1296 (1995).
45. McKhann, G. *et al.* Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* **34,** 939–944 (1984).
46. Lannfelt, L. *et al.* Amyloid precursor protein mutation causes Alzheimer's disease in a Swedish family. *Neurosci. Lett.* **168,** 254–256 (1994).
47. Lannfelt, L. *et al.* Amyloid β-peptide in cerebrospinal fluid in individuals with the Swedish Alzheimer amyloid precursor protein mutation. *Neurosci. Lett.* **199,** 203–206 (1995).
48. Braak, H. & Braak, E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol.* **82,** 239–259 (1991).
49. Mirra, S. S. *et al.* The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology* **41,** 479–486 (1991).
50. Hyman, B. T. *et al.* National Institute on Aging-Alzheime's Association guidelines for the neuropathologic assessment of Alzheimer's disease. *Alzheimers Dement.* **8,** 1–13 (2012).
51. Liao, L. *et al.* Proteomic characterization of postmortem amyloid plaques isolated by laser capture microdissection. *J. Biol. Chem.* **279,** 37061–37068 (2004).
52. Houeland, G. *et al.* Transgenic mice with chronic NGF deprivation and Alzheimer's disease-like pathology display hippocampal region-specific impairments in short- and long-term plasticities. *J. Neurosci.* **30,** 13089–13094 (2010).
53. Townsend, M., Shankar, G. M., Mehta, T., Walsh, D. M. & Selkoe, D. J. Effects of secreted oligomers of amyloid beta-protein on hippocampal synaptic plasticity: a potent role for trimers. *J. Physiol. (Lond.)* **572,** 477–492 (2006).
54. Stosiek, C., Garaschuk, O., Holthoff, K. & Konnerth, A. *In vivo* two-photon calcium imaging of neuronal networks. *Proc. Natl Acad. Sci. USA* **100,** 7319–7324 (2003).

Article 5

The impact II, a very high resolution quadrupole time-of-flight instrument for deep shotgun proteomics

Hybrid quadrupole time-of-flight (QTOF) mass spectrometers are one of the two major principles in mass spectrometry. Over the last decades, QTOF mass spectrometers have greatly evolved in terms of achievable resolution, mass accuracy and dynamic range. The Bruker impact platform of QTOF instruments takes advantage of these developments and here we develop, optimized and evaluate the impact II for shotgun proteomics applications. This instrument features highly efficient ion optics and improvements on the reflectron and ion detector, providing almost twice the resolution compared to its predecessor. These improvements, combined with the principle advantages of TOF technology including high sequencing speed and dynamic range, the m/z independent resolution and comparably low price, make QTOF very appealing for shotgun proteomics. We integrated the impact II into the state of the art proteomics workflow. We then benchmarked this proteomic pipeline for the analysis of complex biological samples. With the QTOF we identified over 4,800 proteins in Hela within 90 min measuring time. Using high pH fractionation, we could quantify over 11,000 proteins in murine cerebellum, which is the deepest proteome acquired with a QTOF instrument to date.

Molecular & Cellular Proteomics

Beck S.; Michalski A.; Raether O.; Lubeck M.; Kaspar S.; Goedecke N.; Baessmann C.; **Hornburg D.;** Meier F.; Paron I.; Kulak N.; Cox J. and Mann M.,

Contribution: This paper is a collaborative effort between the group of Mathias Mann (MPI Biochemistry) and Bruker. To benchmark and optimize the performance of the QTOF in complete proteome analyses of complex biological samples, I prepared neuronal cell lines and control cell lines. Furthermore, I contributed to the data analysis and interpretation of the cell line and the murine cerebellum proteomics results.

The Impact II, a Very High-Resolution Quadrupole Time-of-Flight Instrument (QTOF) for Deep Shotgun Proteomics*⑤

**Scarlet Beck‡, Annette Michalski§, Oliver Raether§, Markus Lubeck§,
Stephanie Kaspar§, Niels Goedecke§, Carsten Baessmann§, Daniel Hornburg‡,
Florian Meier‡, Igor Paron‡, Nils A. Kulak‡, Juergen Cox¶, and Matthias Mann‡‖**

Hybrid quadrupole time-of-flight (QTOF) mass spectrometry is one of the two major principles used in proteomics. Although based on simple fundamentals, it has over the last decades greatly evolved in terms of achievable resolution, mass accuracy, and dynamic range. The Bruker impact platform of QTOF instruments takes advantage of these developments and here we develop and evaluate the impact II for shotgun proteomics applications. Adaption of our heated liquid chromatography system achieved very narrow peptide elution peaks. The impact II is equipped with a new collision cell with both axial and radial ion ejection, more than doubling ion extraction at high tandem MS frequencies. The new reflectron and detector improve resolving power compared with the previous model up to 80%, *i.e.* to 40,000 at *m/z* 1222. We analyzed the ion current from the inlet capillary and found very high transmission (>80%) up to the collision cell. Simulation and measurement indicated 60% transfer into the flight tube. We adapted MaxQuant for QTOF data, improving absolute average mass deviations to better than 1.45 ppm. More than 4800 proteins can be identified in a single run of HeLa digest in a 90 min gradient. The workflow achieved high technical reproducibility (R2 > 0.99) and accurate fold change determination in spike-in experiments in complex mixtures. Using label-free quantification we rapidly quantified haploid against diploid yeast and characterized overall proteome differences in mouse cell lines originating from different tissues. Finally, after high pH reversed-phase fractionation we identified 9515 proteins in a triplicate measurement of HeLa peptide mixture and 11,257 proteins in single measurements of cerebellum—the highest proteome coverage reported with a QTOF instrument so far.   *Molecular & Cellular Proteomics 14: 10.1074/mcp.M114.047407, 2014–2029, 2015.*

Building on the fundamental advance of the soft ionization techniques electrospray ionization and matrix-assisted laser desorption/ionization (1, 2), MS-based proteomics has advanced tremendously over the last two decades (3–6). Bottom-up, shotgun proteomics is usually performed in a liquid chromatography-tandem MS (LC-MS/MS)[1] format, where nanoscale liquid chromatography is coupled through electrospray ionization to an instrument capable of measuring a mass spectrum and fragmenting the recognized precursor peaks on the chromatographic time scale. Fundamental challenges of shotgun proteomics include the very large numbers of peptides that elute over relatively short periods and peptide abundances that vary by many orders of magnitude. Developments in mass spectrometers toward higher sensitivity, sequencing speed, and resolution were needed and helped to address these critical challenges (7, 8). Especially the introduction of the Orbitrap mass analyzers has advanced the state of the art of the field because of their very high resolution and mass accuracy (9, 10). A popular configuration couples a quadrupole mass filter for precursor selection to the Orbitrap analyzer in a compact benchtop format (11–13).

In addition to the improvements in MS instrumentation, there have been key advances in the entire proteomics workflow, from sample preparation through improved LC systems and in computational proteomics (14–16). Together, such

---

[1] The abbreviations used are: LC-MS/MS, liquid chromatography-tandem MS; CAA, chloroacetamide; ES, electrospray; FDR, false discovery rate; Hepa 1–6, mouse hepatoma; ID, inner diameter; LFQ, label-free quantification; MCP, multichannel plate; MEFs, mouse embryonic fibroblasts; MS/MS, tandem mass spectrometry; NSC-34, spinal cord neuron-neuroblastoma; QTOF, quadrupole time-of-flight; TOF, time-of-flight; UPS, universal protein standard.

advances are making shotgun proteomics increasingly comprehensive and deep analyses can now be performed in a reasonable time (13, 17–19). Nevertheless, complete analysis of all expressed proteins in a complex system remains extremely challenging and complete measurement of all the peptides produced in shotgun proteomics may not even be possible in principle (20, 21). Therefore, an urgent need for continued improvements in proteomics technology remains.

Besides the Orbitrap analyzer and other ion trap technologies, the main alternative MS technology is time-of-flight, a technology that has been used for many decades in diverse fields. The configuration employed in proteomics laboratories combines a quadrupole mass filter via a collision cell and orthogonal acceleration unit to a reflectron and a multichannel plate (MCP) detector (22). TOF scans are generated in much less than a millisecond (ms), and a number of these "pulses" are added to obtain an MS or MS/MS spectrum with the desired signal to noise ratio. Our own laboratory has used such a quadrupole time-of-flight (QTOF) instrument as the main workhorse in proteomics for many years, but then switched to high-resolution trapping instruments because of their superior resolution and mass accuracy. However, TOF technology has fundamental attractions, such as the extremely high scan speed and the absence of space charge, which limits the number of usable ions in all trapping instruments. In principle, the high spectra rate makes TOF instruments capable of making use of the majority of ions, thus promising optimal sensitivity, dynamic range and hence quantification. It also means that TOF can naturally be interfaced with ion mobility devices, which typically separate ions on the ms time scale. Data independent analysis strategies such as MS^E, in which all precursors are fragmented simultaneously (23, 24) or SWATH, in which the precursor ion window is rapidly cycled through the entire mass range (25), also make use of the high scanning speed offered by QTOF instruments. It appears that QTOFs are set to make a comeback in proteomics with recent examples showing impressive depth of coverage of complex proteomes. For instance, using a variant of the MS^E method, identification of 5468 proteins was reported in HeLa cells in single shots and small sample amounts (26). In another report, employing ion mobility for better transmission of fragment ions to the detector led to the identification of up to 7548 proteins in human ovary tissue (27).

In this paper, we describe the impact II™, a benchtop QTOF instrument from Bruker Daltonics, and its use in shotgun proteomics. This QTOF instrument is a member of an instrument family first introduced in 2008, which consists of the compact, the impact, and the maXis. The original impact was introduced in 2011 and was followed by the impact HD, which was equipped with a better digitizer, expanding the dynamic range of the detector. With the impact II, which became commercially available in 2014, we aimed to achieve a resolution and sequencing speed adequate for demanding shotgun proteomics experiments. To achieve this we developed

an improved collision cell, orthogonal accelerator scheme, reflectron, and detector. Here we measure ion transmission characteristics of this instrument and the actually realized resolution and mass accuracy in typical proteomics experiments. Furthermore, we investigated the attainable proteome coverage in single shot analysis and we ask if QTOF performance is now sufficient for very deep characterization of complex cell line and tissue proteomes.

EXPERIMENTAL PROCEDURES

Preparation of HeLa Lysates—HeLa cells (ATCC, S3 subclone) were cultured in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal bovine serum, 20 mM glutamine and 1% penicillin-streptomycin (all from PAA Laboratories, Freiburg, Germany). Cells were collected by centrifugation at 200 × g for 10 min, washed once with cold phosphate buffered saline (PBS) and centrifuged again. Supernatant was carefully discarded and the cell pellet shock frozen in liquid nitrogen and stored at −80 °C until further use. A pellet containing 5 × 10^7 cells was resuspended in 1.5 ml of ice cold Milli-Q water, then an equal volume of trifluoroethanol (Sigma-Aldrich, Taufkirchen, Germany) was added. The cell suspension was kept on ice for 10 min, vortexed for 1 min and sonicated for 2 min at 20% duty cycle and output control 3 (Branson Ultrasonics sonifier, Danbury, CT; model 250). After the addition of 200 µl Tris (pH 8.5, final concentration: 100 mM), 400 µl TCEP (final concentration: 10 mM) and 400 µl 2-chloroacetamide (CAA) (final concentration: 40 mM) the lysate was incubated for 10 min at 95 °C. Then the sample was diluted to 15 ml with 50 mM ammonium bicarbonate. The mixture was digested by adding LysC (Wako Chemicals GmbH, Neuss, Germany; ratio 1 µg LysC:100 µg sample protein) for 2 h at 37 °C, followed by adding trypsin (ratio 1 µg trypsin:75 µg sample protein, Promega GmbH, Mannheim, Germany) at 37 °C overnight. After a further digestion with trypsin (ratio 1:125) for 5 h at 37 °C, the digested peptides with an estimated concentration of 1 µg/µl were diluted 1:4 with water and acidified by adding formic acid (FA) (final concentration: 0.2%) and purified on Sep-Pak tC18 cartridges (Waters, Milford, MA) according to manufacturer's instructions. Peptide concentration was determined using a NanoDrop spectrophotometer (Thermo Scientific, Wilmington, DE).

Preparation of Yeast Lysates—Saccharomyces cerevisiae strains BY4742 and BY4743 (EUROSCARF) were grown at 30 °C in yeast extract peptone dextrose (YPD) media (10 g/l BactoYeast extract, 20 g/l Bacto™ peptone (BD), 2% w/v glucose). Cells were grown to log phase (OD$_{600}$ of 0.6), harvested by centrifugation at 1600 × g for 10 min at 4 °C, washed with cold Milli-Q water and then collected again by centrifugation at 10,000 × g for 5 min at 4 °C. Cells were lysed in 1% sodium deoxycholate, 10 mM TCEP, 40 mM CAA in 100 mM Tris pH 8.5, boiled for 10 min at 95 °C and sonicated for 3 min at 30% duty cycle and output control 3 (Branson Ultrasonics sonifier; model 250). Protein concentrations were determined by tryptophan fluorescence emission assay. Cell lysates were diluted 1:2 with Milli-Q water and digested by adding LysC (Wako Chemicals GmbH, ratio 1 µg LysC:50 µg sample protein) for 4 h at 37 °C, followed by adding again LysC (ratio 1:50) overnight at 37 °C. An equal volume of ethyl acetate acidified with 1% TFA was added to the solution, samples were vortexed for 2 min and digested peptides were purified with SDB-RPS StageTips as described in Kulak *et al.* (19). Peptide concentrations were determined using a NanoDrop spectrophotometer.

Preparation of MEFs, Hepa, and NSC Cell Line Lysates—Spinal cord neuron-neuroblastoma (NSC-34) (CED-CLU140, Biozol, Eching, Germany), mouse embryonic fibroblasts (MEFs) (American Type Culture Collection, Manassas, VA), and mouse hepatoma (liver cancer, Hepa 1–6) (CRL-1830, American Type Culture Collection) cell lines

were cultured and proteins prepared as previously described (28). Briefly, the cells were lysed in lysis buffer (4% SDS, 10 mM Hepes, pH 8.0) during sonication for 15 min (level 5, Bioruptor; Diagenode, Seraing (Ougrée) - Belgium). Cell lysis was followed by reduction of disulfide bonds with 10 mM DTT for 30 min and subsequent alkylation with 55 mM IAA for 45 min. To remove the detergent, cold acetone (−20 °C) was added to 100 μg of proteins to a final concentration of 80% v/v, and proteins were precipitated for at least 2 h at −20 °C. The suspension was centrifuged for 15 min (4 °C, 16,000 × g) and the precipitate was washed with 80% acetone (−20 °C) prior to re-suspension in 50 μl of 6 M urea/2 M thiourea, 10 mM Hepes, pH 8.0. An initial digestion step (3 h) was carried out after the addition of 1 μg of LysC, followed by dilution with four volumes of 50 mM ammonium bicarbonate and the final digestion with 1 μg of trypsin overnight at room temperature. The resulting peptide mixtures were desalted on SDB-RPS StageTips (29) and subjected to single shot LC-MS/MS analysis.

Preparation of Cerebellum Lysates—Cerebellum from a single mouse (strain: C57Bl6) was homogenized in 4% SDS in 100 mM Tris pH 7.6 using a FastPrep 24 homogenizer (MP Biomedicals, Eschwege, Germany), incubated for 10 min at 95 °C and sonicated for 3 min at 30% duty cycle and output control 3 (Branson Ultrasonics sonifier; model 250). To remove the detergent, acetone (−20 °C) was added to a final concentration of 80% v/v and proteins were precipitated overnight at −20 °C. Supernatants were carefully discarded after centrifugation at 1600 × g for 20 min at 4 °C, and the pellets were washed with 80% acetone (−20 °C). The protein pellets were dissolved in 8 M Urea in 10 mM Hepes and protein concentrations were determined by the tryptophan fluorescence emission at 350 nm using an excitation wavelength of 295 nm. Proteins were reduced with 10 mM DTT for 30 min and alkylated with 55 mM iodoacetamide for 20 min. After addition of thiourea to a final concentration of 0.1 M, samples were digested by adding LysC (Wako Chemicals, ratio 1 μg LysC:100 μg sample protein) for 3 h at RT, diluted with four volumes of 50 mM ammonium bicarbonate, and further digested with trypsin (ratio 1 μg trypsin:100 μg sample protein, Promega) at RT overnight. After a further digestion with LysC and trypsin (ratio 1:100) for 8 h at RT, digested peptides were acidified by adding TFA (final concentration: 0.5%) and purified on Sep-Pak tC18 cartridges (Waters) according to manufacturer's instructions. Peptide concentrations were determined using a NanoDrop spectrophotometer.

Sample Preparation for Quantification—Universal Proteomics Standard (UPS-1, Sigma-Aldrich) and Proteomics Dynamic Range Standard (UPS-2, Sigma-Aldrich), both containing 48 human proteins, either at equimolar concentrations (UPS-1) or formulated into a dynamic range of concentrations, covering five orders of magnitude (UPS-2), were prepared according to ref (30). Predigested yeast sample (Promega) was re-suspended in 0.1% trifluoroacetic acid to a final concentration of 500 ng/μl. Digested UPS-2 sample was spiked in two different amounts of 250 fmol to 2.5 amol peptide amount for sample 1 and 500 fmol to 5 amol for sample 2 into 500 ng yeast background, thereby creating two samples with a theoretical ratio 1:1 for the yeast proteome and 1:2 for the UPS peptides. In another sample, digested UPS-1 sample (25 fmol for all components) was spiked into 500 ng yeast.

High-pH Reverse-Phase Fractionation—We performed high-pH reversed-phase peptide prefractionation with fraction concatenation on 175 μg HeLa or cerebellum peptides on a 2.1 × 300 mm Acquity UPLC Peptide BEH column packed with 130 Å pore, 1.7 μm particle size C_{18} beads (Part No. 186005792, Waters). A gradient of basic reversed-phase buffers (Buffer A: 0.1% formic acid, ammonium hydroxide pH 10; Buffer B: 0.1% formic acid, 80% acetonitrile, ammonium hydroxide pH 10) was run on a Prominence HPLC system (Shimadzu, Duisburg, Germany) at a flow rate of 150 μl/min at 60 °C.

The LC run lasted for 240 min with a starting concentration of 5% buffer B increasing to 30% over the initial 120 min and a further increase in concentration to 60% over 70 min. This elution gradient was followed by a 95% wash and re-equilibration. Fraction collection started after 0.2 ml elution and fractions were collected every 140 s resulting in 72 fractions used for concatenation into 24 fractions as described previously (31).

Inlet Capillary and CaptiveSpray—In our instrument, in contrast to many other commercial ion source designs, the high voltage for the electrospray (ES) process is applied to the vacuum capillary inlet, whereas the sprayer is kept at ground, which allows for a simpler source design (supplemental Fig. S1*A*). To electrically decouple the ES voltage and the electrical potential of the vacuum section, we use an inlet capillary made from high resistive glass (~1GOhm). Positioning the ES voltage at the capillary entrance means that the ions are transported opposite to the electrical gradient by the gas flow (32). In this configuration, charged molecules travel somewhat slower than the surrounding gas. According to Bernoulli's law, ions are then focused toward the area of highest gas velocity along the center axis of the capillary. The set-up tends to reduce the contamination of the inner capillary walls (33).

The Bruker CaptiveSpray nanoflow ES source is directly attached to the vacuum inlet capillary via a short capillary extension that can be heated using the instrument's drying gas (Supplemental Fig. S1*A*). The spray tip is automatically mechanically aligned on axis with the capillary inlet without the need for any adjustments. The principle of the CaptiveSpray is a vortex gas (usually air) that sweeps around the emitter spray tip at three different stages. The first one is designed to assist spray formation, the second and third one help to focus the spray plume into the MS inlet capillary. All three flows are created solely by the vacuum of the MS system, which requires that the entire source is vacuum sealed.

The spray emitter consists of a 2 cm long, 20 μm ID fused silica capillary. Its tip is etch-tapered, thus the inner diameter remains constant to the very end of the tip making it very robust against clogging. Furthermore, it also allows using the same emitter at flow rates ranging from 50 nl/min to 5 μl/min, thereby supporting a wide range of column types. Fused silica columns, which are often used for proteomics, are typically connected to the emitter via a low dead volume union (supplemental Fig. S1*A*), which also provides the electrical contact for keeping the electrospray at ground potential.

Minimizing Postcolumn Dead Volume—Using the described design, the CaptiveSpray source provides very stable ionization; however, when we initially coupled it to the LC set-up used in the Munich laboratory (17, 34), we observed broader LC peak elution distributions than we normally do (supplemental Fig. S1*B*). Furthermore, we wished to incorporate a column oven and pulled tip columns. We therefore constructed a modified source, which keeps the back end of the CaptiveSpray but replaces the front end by the standard set-up used in our department. The modified set-up incorporating the tip column is displayed in supplemental Fig. S1*C*. The modified design of the column holder allows for exact aligning and fixation of the column inside the CaptiveSpray source. Electrical grounding was applied using a connecting tee at the column head. This setup produced the desired, narrow LC peak distributions (supplemental Fig. S1*D*) and was used for the proteomic analyses described in this article.

LC-MS/MS Analysis—We used an Easy nLC-1000 (Thermo Fisher Scientific) on-line coupled to an impact II (Bruker Daltonics) with a CaptiveSpray ion source (Bruker Daltonics). The peptide mixtures (1 μg) were loaded onto an in-house packed column (50 cm, 75 μm inner diameter) filled with C_{18} material (ReproSil-Pur C_{18} AQ 1.9 μm reversed phase resin, Dr. Maisch GmbH, Ammerbuch-Entringen, Germany). Chromatographic separation was carried out using a linear gradient of 5–30% buffer B (80% ACN and 0.1% FA) at a flow rate of

250 nl/min over 90 min. Because of loading and washing steps, the total time for an LC-MS/MS run was about 40 to 50 min longer.

Generally, LC-MS/MS data were acquired using a data-dependent auto-MS/MS method selecting the 17 most abundant precursor ions in cycle for fragmentation and an MS/MS summation time adjusted to the precursor intensity (Compass 1.8 acquisition and processing software, Bruker Daltonics). For the deep proteome measurements of a cell line in combination with peptide fractionation, we used a "dynamic method," with a fixed cycle time of 3 s. The mass range of the MS scan was set to extend from m/z 150 to 1750. Dynamic exclusion duration was 0.4 min. Isolation of precursor ions was performed using an m/z dependent isolation window of 1.5–5 Th. The collision energy was adjusted between 23–65 eV as a function of the m/z value.

For the quantitative analysis of the UPS standards in yeast we used a trapping column set-up (PepMap pre-column, 2 cm x 100 μm; Thermo Scientific) and a Dionex HPLC pump (Ultimate 3000, Thermo Scientific). For this experiment, peptides were separated on a PepMap UHPLC column (50 cm x 75 μm, 2 μm particles; Thermo Scientific) using a 90 min multistep ACN gradient (buffer A: 0.1% FA; buffer B 100% ACN in 0.1% FA). The unmodified CaptiveSpray ion source (see above) was used to interface the LC system to the impact II. For quantification full scan MS spectra were acquired at a spectra rate of 1Hz followed by acquisition of 1 MS/MS spectrum. Six replicates per sample were acquired. For data acquisition of the UPS-1 in yeast sample, the 17 most intense precursor ions were selected for fragmentation, resulting in a total cycle time of 1.2 s.

Intact Protein Analysis—Adalimumab was cleaved at the hinge region with IdeS (FabRICATOR, Genovis) and reduced to obtain the Fc/2, Fd and light chain sub units as recently described in (35). The subunits were separated by chromatography (35) prior to analysis on the impact II. Data was analyzed using the SNAP algorithm to fit the theoretical pattern (36, 37).

Development of MaxQuant for QTOF Data—In general all processing steps from the standard MaxQuant computational workflow, which was optimized for the analysis of Orbitrap data, are also applied to QTOF data. The nonlinear mass recalibration algorithm experienced major adaptations. Its original form for the Orbitrap applies a recalibration function with nonlinear dependence on the two variables m/z and retention time. It has been extended to include the peak intensity as a third dimension that the mass recalibration depends on. This is necessary because of appreciable systematic nonlinear intensity dependent peak mass shifts that are typically found in time of flight data. The intensity dependence is parameterized as a polynomial in the logarithm of peak intensities. The new mass recalibration algorithm allows for high mass accuracy without the use of internal or external calibrants.

We added a new instrument type called "Bruker QTOF" in which several relevant parameters of algorithms for the processing of spectra are set to default values that are suitable for the analysis of data generated by the impact family. These parameters include mass matching windows for the assembly of 3D peaks, mass tolerances for assembling isotope patterns and labeling pairs, initial peptide mass tolerance windows for the Andromeda search and minimum required number of scans per 3D peak. Raw data can be immediately read from the proprietary Bruker binary format and no conversion to intermediate file formats is needed. Peak centroids are utilized as determined by the centroiding algorithms of in the Bruker software. The viewer module of MaxQuant is enabled for QTOF data, among other features allowing to visualize MS data in m/z-retention time maps and to annotate and export MS/MS spectra to fulfil journal requirements for reporting of spectral evidence.

Analysis of Proteomic Data—All data were analyzed with the MaxQuant software (version 1.5.2.8 or version 1.5.0.1) (38, 39) with the Andromeda search engine (38) with the adaptions and developments described above. The false discovery rate (FDR) was set to 1% for both proteins and peptides and we specified a minimum length of seven amino acids. MaxQuant scored peptides for identification based on a search with an initial allowed mass deviation of the precursor ion of up to 0.07 Da after time-dependent recalibration of the precursor masses. The allowed fragment mass deviation was 40 ppm. The Andromeda search engine was used for the MS/MS spectra search against the Uniprot human database (downloaded on June 21, 2014, containing 88,976 entries and 247 contaminants), the Uniprot Saccharomyces cerevisiae database (downloaded on June 21, 2014, containing 6643 entries), the Uniprot mouse database (downloaded on June 21, 2014, containing 51,573 entries) and UPS fasta file provided by Sigma-Aldrich (http://www.sigmaaldrich.com/life-science/proteomics/mass-spectrometry/ups1-and-ups2-proteomic.html) for quantitative study. Enzyme specificity was set as C-terminal to Arg and Lys, also allowing cleavage at proline bonds and a maximum of two missed cleavages. Carbamidomethylation of cysteine was selected as fixed modification and N-terminal protein acetylation and methionine oxidation as variable modifications.

The "match between runs" feature of MaxQuant was used to transfer identifications to other LC-MS/MS runs based on their masses and retention time (maximum deviation 0.7 min) and this was also used in quantification experiments. Quantifications were performed with the label-free algorithms described recently (39). We required a minimum peptide ratio count of two and at least one "razor peptide" for quantification. For cerebellum, the quantification was based on normalized protein intensities. Further analysis of data was performed in the MaxQuant Viewer, in the Perseus post data acquisition package that is part of MaxQuant (all freely available at www.maxquant.org) and in the R statistical computing environment (40).

Potential contaminants as well as proteins identified only by site modification were strictly excluded from further analysis.

For the quantitative analysis of the UPS standards in yeast, entries were only accepted if they had valid values in all 12 replicates. Results were then filtered for Welch-significant regulation of UPS-2 proteins.

Analysis of the yeast samples were based on label-free intensities (LFQ values). After filtering (3 valid values in at least one group), remaining missing values were imputed from a normal distribution (width: 0.3; down shift: 1.8). Two-sample t test was performed with a FDR < 0.01.

For global cell line comparison, triplicates were analyzed twice for a total of six single shot measurements per cell line (except the NSC-34 cell line, which was only measured once). For the principal component analysis (PCA) of the different cell lines, we furthermore limited the data set (LFQ intensities) to entries with a minimum of four valid values in at least one group of six replicates. Remaining missing values were imputed from a normal distribution (see above).

Protein intensity (summed peptide intensity) for cerebellum samples were divided by the molecular weight for ranking by proteins abundance. Annotations (GO molecular function, biological process, cellular component; KEGG and Uniprot Keywords) were matched to protein groups with Perseus. We performed a 1D annotation enrichment (41) on the normalized protein intensities. To evaluate the proteome coverage, we counted the occurrence of categories in our sample and compared it to the category count for the complete murine proteome in Perseus.

MS raw data and data for protein and peptide identification and quantification were submitted as supplementary tables to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD001592.
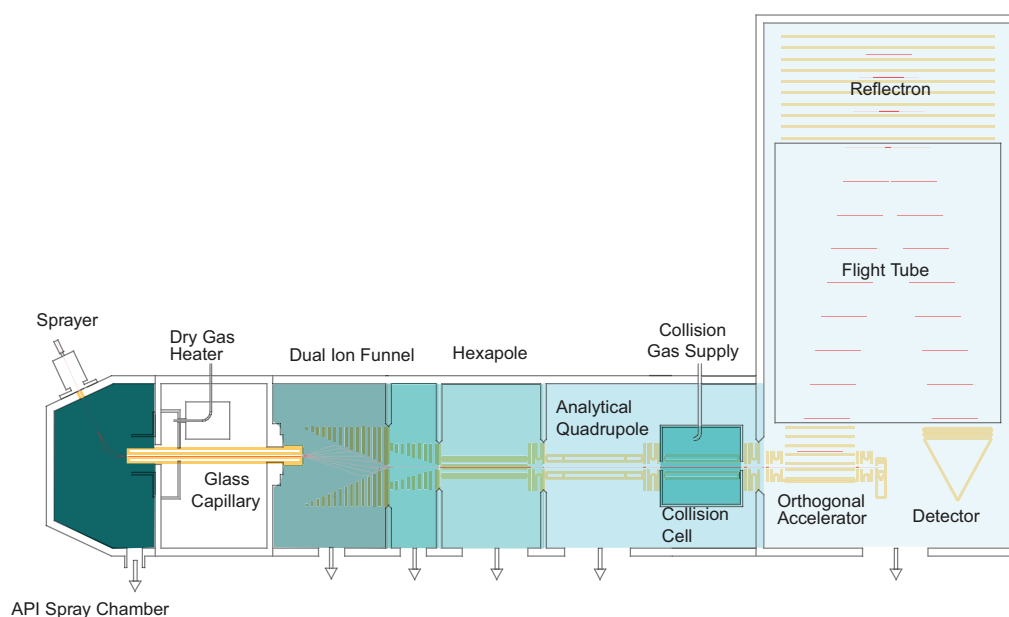
FIG. 1. **Schematic of the impact II mass spectrometer (not to scale).**

RESULTS AND DISCUSSION

Overview of the Instrument—The Bruker impact II is a QTOF in a benchtop format, featuring several improvements in its design (Fig. 1). Briefly, ions are produced in the CaptiveSpray, which is in an encased nanoelectrospray source that features a well-defined gas stream to guide the ions into the vacuum via a capillary inlet. A double ion funnel, based on principles described by Smith and co-workers (42), is positioned off axis, which prevents neutrals from further transmission along the ion path. The pressure drops by several orders of magnitude from the capillary exit to the postfunnel stage (3 mbar to 3×10^{-4} mbar), while the ion current is virtually undiminished (see below). Additionally, the funnel allows for soft transfer based on low electrical field strength independent of the mass (typically 10 V/cm, much lower than in nozzle-skimmer designs). By introducing electrical acceleration in-between the two funnels, in-source fragmentation can still be achieved intentionally. There is a hexapole ion guide between funnel and the analytical quadrupole mass filter, which has a monolithic design based on high precision glass. Precursor ions can be isolated by this quadrupole for subsequent fragmentation in the collision cell. Intact ions or fragments can be stored and extracted from the collision cell and enter the orthogonal deflection region as a very narrowly focused ion beam ($< 500 \ \mu$m). Here they are accelerated into a field-free drift region. A newly designed, two-stage reflectron further compensates the velocity distribution orthogonal to the beam direction. Finally, the ions impinge on an MCP detector coupled to a 10-bit, very high frequency (50 Gbit/s), zero noise digitizer. Data collection is coordinated by the Bruker Compass data

system and in the experiments described here, post-acquisition data processing is performed in the MaxQuant environment.

Optimization of the Collision Cell—Efficient fragmentation of precursor ions on an LC-MS/MS time scale is a key for the identification of peptides in shotgun proteomics strategies. We optimized several aspects of the collision cell (supplemental Fig. S2A): Precise geometrical alignment allows focusing of the ions along the axis of the collision cell, directly translates into well-defined starting conditions for the orthogonal accelerator and is therefore mandatory for high mass resolution. This is implemented via a quadrupolar configuration of the collision cell device providing a narrow pseudo potential well (43). We also introduced a radial ejection step between any two MS or MS/MS experiments, in order to reduce the dead time. The ion path has to be emptied to avoid crosstalk between two consecutive spectra without introducing substantial ion losses. Most importantly, we optimized the time of ion fragmentation and extraction within the fragment spectra to ensure efficient high frequency MS/MS by implementing an electrical axial field gradient. This gradient directs the ions toward the exit of the collision cell, reducing the time it takes for the first ions to reach the extraction lens after quenching the collision cell. From here, they can be rapidly released toward the orthogonal accelerator, forming packages that match the orthogonal pulser frequency. The ion densities after 3 ms (upper red traces in supplemental Fig. S2B–S2C) reveals two important aspects: Using the axial field gradient results in comparable ion density at the collision cell exit in much shorter time than without, *i.e.* after about 1 ms instead of
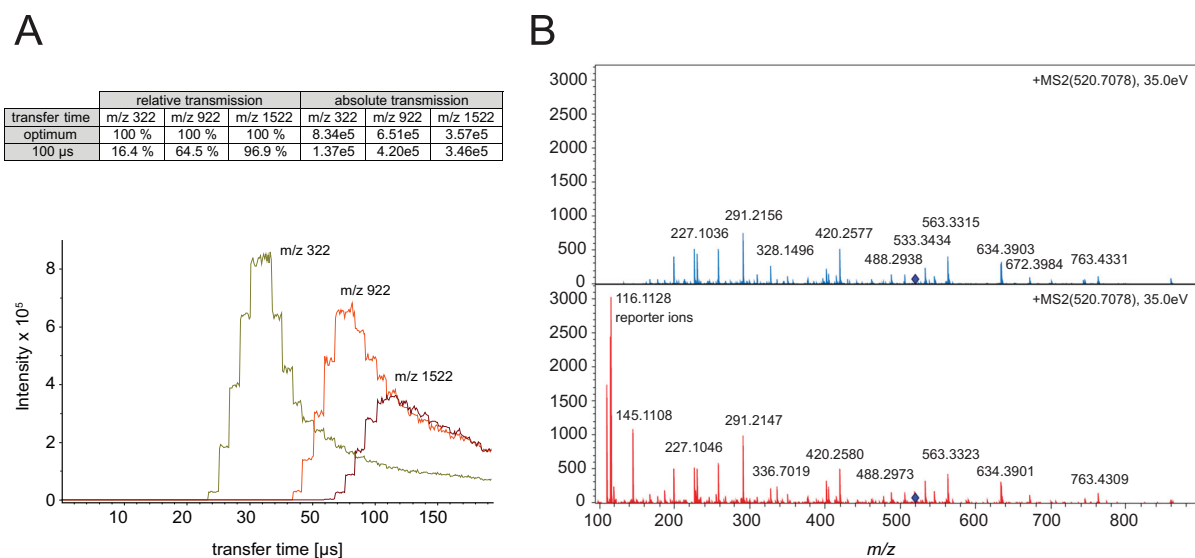
A

transfer time	relative transmission			absolute transmission		
	m/z 322	m/z 922	m/z 1522	m/z 322	m/z 922	m/z 1522
optimum	100 %	100 %	100 %	8.34e5	6.51e5	3.57e5
100 μs	16.4 %	64.5 %	96.9 %	1.37e5	4.20e5	3.46e5

B

FIG. 2. *A, Absolute ion intensities of m/z 322, 922 and 1522 as function of the transfer time.* The maxima of the distributions for each *m/z* value are the optimal times for efficient transfer into the flight tube. *B*, Fragment spectrum of the iTRAQ labeled peptide LFTGHPETLEK without (blue) and with (red) transfer time stepping that adds the lower *m/z* range including the reporter ions without sacrificing intensity in the higher mass range.

3 ms, respectively. This suggests a three times faster ion transfer. Moreover, the overall number of ions within the collision cell is reduced to less than 50% at the moment of quenching, which should reduce ion losses related to quenching accordingly.

The pseudo potential calculations and simulation were confirmed in different experiments comparing the performance of the optimized collision cell on the fragmentation yield of Glu-Fibrino-Peptide B at different MS/MS acquisition rates (supplemental Fig. S2D). This revealed that the reduced quench losses in fact improve the number of ions detected at a spectra rate of 16 Hz by a factor of two. We further observed that the axial field gradient improves the stability of the system even in the presence of slight contaminations on the rods.

High Transfer Efficiency to the Orthogonal Acceleration Unit—The ions travel through the flight tube and require as much time as the largest *m/z* species needs to reach the detector, before the HV pulser can send the next ion package toward the detector (typically between 100 and 150 μs). To avoid excessive loss of ions, orthogonal TOF instruments are therefore often operated in a mode in which the ions are stored in the collision cell during the TOF scan and released in time for the next extraction pulse of the orthogonal accelerator. This would allow for 100% duty cycle if all ions were indeed transferred such that they arrive in the orthogonal accelerator at the same time and with the same kinetic energy. In practice, however, the extraction time from the collision cell toward the gate lens is a function of ion mobility. We have analyzed these combined effects by varying the time from opening the gate lens to the extraction pulse of the

orthogonal accelerator (transfer time) for different *m/z* ratios (Fig. 2A). Simulations of ion trajectories and extraction times reveal that about 80% of a single ion species can be accelerated into the drift tube of the TOF analyzer under optimal transfer time conditions. In the impact II the high transfer efficiency is further optimized by reducing the distance between the trapping region and the orthogonal accelerator—it is about four times higher compared with conventional QTOF systems operated in continuous operation mode. Figure 2A summarizes the relative transmission efficiency of selected precursors (*m/z* = 322 Th, 922 Th and 1522 Th) at a transfer time of 100 μs, which is a standard setting to cover the mass range relevant in shotgun proteomics. The analysis highlights that ions with low *m/z* are compromised most, which we have previously counteracted in proteomics experiments by adding special extraction conditions (44). To tackle this problem in a more general way, Bruker introduced the "transfer time stepping" operation mode, where first high mobility species are extracted followed by species with a lower mobility. During the initially short opening times of the gate (typically 50% of the total scan) only the higher mobility, low *m/z* ions pass the gate lens, while the lower mobility ions are still accumulated in the collision cell. In the second transfer time steps, extraction times are increased to allow the low mobility, higher *m/z* ions to pass the gate lens. Spectra with and without transfer time stepping, reveal the beneficial effect on low mass ions without appreciable loss in the standard mass range (Fig. 2B). This is particularly beneficial for labeling experiments that rely on quantification of low mass reporter ions as shown in the figure. These ions are transferred with greater than 60% effi-
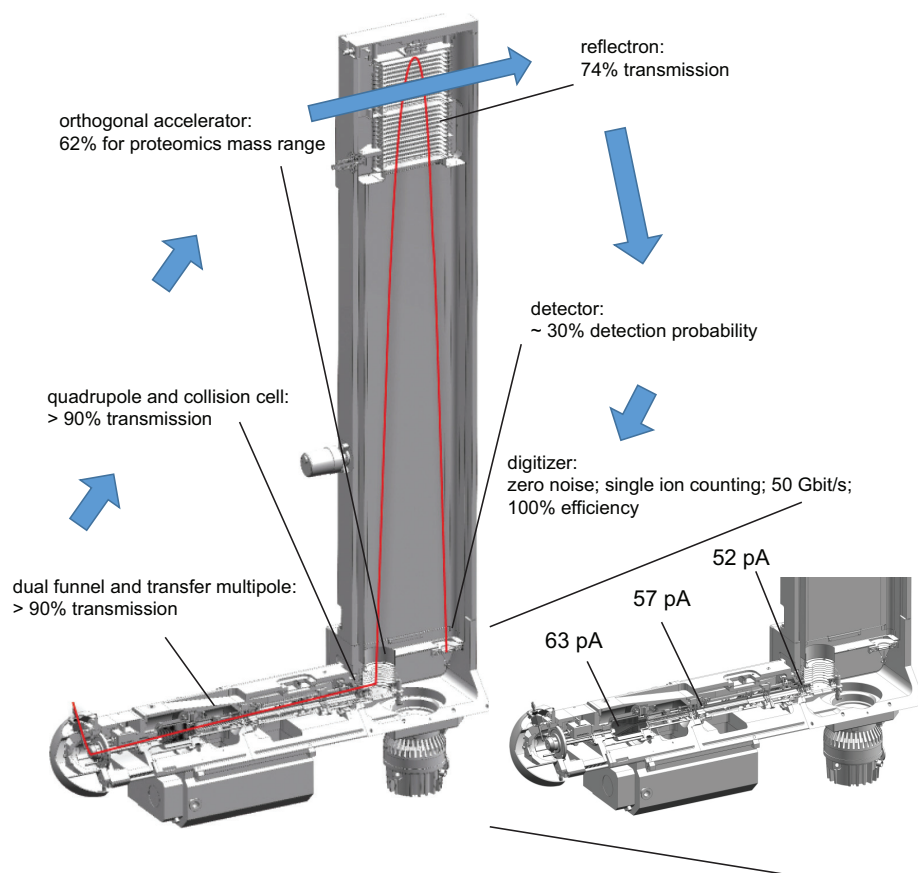
Fig. 3. **Ion transfer efficiency of mass range most relevant in shotgun-proteomics experiments (*m/z* 100–1500).** Inset shows net analyte ion currents at the indicated measurement points and transfer efficiencies at the orthogonal accelerator at various stages along the flight path of the impact II instrument.

ciency during the low *m/z* extraction phase and the total amount of ions in the analyzed fragment spectrum is increased by 58%.

Together, our developments led to an overall transmission efficiency of > 60% into the orthogonal acceleration unit. This compares favorably to a recent report, in which ion mobility was performed on fragment ions and their arrival times were synchronized with orthogonal extraction, which lead to an up to 10-fold improvement of detection sensitivities to standard operation (27).

Sensitivity and Ion Transfer—The number of ions that successfully pass through the instrument and are finally recorded determine a mass spectrometer's sensitivity. We were interested in the transfer efficiencies along the ion path, from entering the vacuum system to the detector. To experimentally determine this, we infused a 1 pmol/μl BSA solution or blank solution and measured the difference in the ion current between these conditions. When operating the outlet of the capillary and the funnel region as a Faraday cage, we measured a net ion current of 63 pA, which we defined as the

starting value (100%) (Fig. 3). The large ion acceptance aperture of the first funnel efficiently captures in the ion flux leaving the capillary. It transfers the ions to the second funnel, which also passes the ions in an almost lossless manner through the next stage as evidenced by a net current reading of 57 pA (>90%) after the hexapole. Likewise, more than 90% of the ions are transferred through the quadrupole (in nonmass selecting operation) and the collision cell. More than 60% of these ions are transmitted into the flight tube by the orthogonal accelerator (see above). The reflectron contains two grids, which are each passed twice, leading to a geometrically defined overall transmission of 74%. Although all these ions hit the MCP, not all of them enter the channels and not all of them result in secondary electrons (detector quantum yield). However, when secondary ions are generated they are greatly amplified ($>10^6$ fold), and can be efficiently discriminated from electronic noise (zero noise detection). Together, this leads to an estimated detection probability of 30% for our MCP detector. Although not all of these measurements and estimates are very precise, they suggest an overall detection
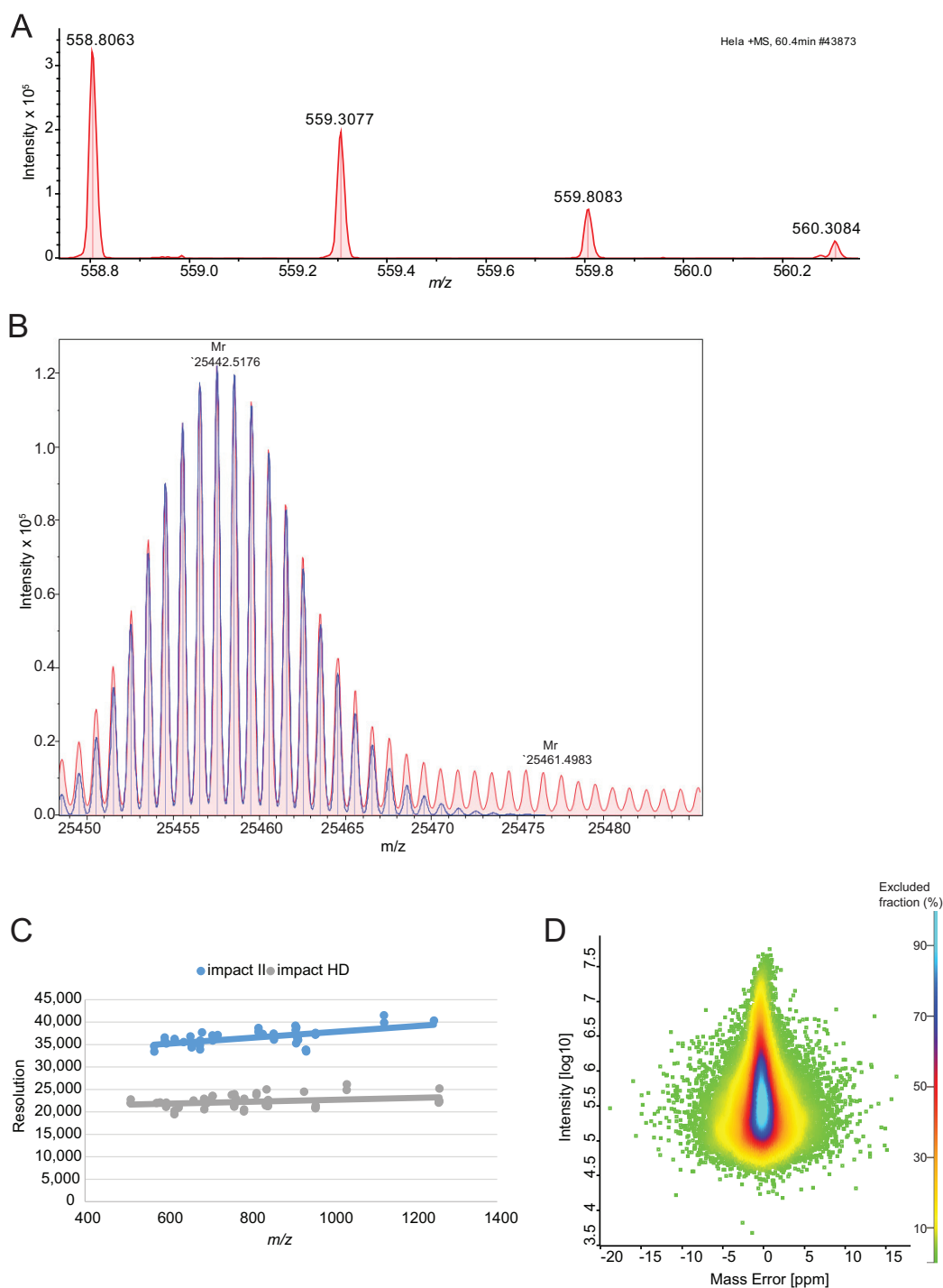
Fig. 4. **Resolution and mass accuracy of *A*, a peptide isotope cluster (*m/z* 558.8063, *r* = 33k) and *B*, Fd unit of Adalimumab (*m/z* 25442.5157, *r* = 63k, 0.26 ppm).** Overall improvement of the resolution with the improved detector, *C*, and the achieved mass accuracy dependent on the summed peptide intensity, *D*, in a shotgun proteomics experiment using the QTOF optimized version of MaxQuant.

TABLE I

Identification from HeLa and yeast lysate triplicate analysis using a standard 90 min gradient

	MS scans	Isotope pattern	MS/MS scans	Identification rate [%]	Peptide sequences identified	Proteins identified
HeLa_1	7002	759,774	79,704	47.09	35,547	4870
HeLa_2	7181	769,355	79,876	47.22	35,572	4864
HeLa_3	7272	796,086	81,389	46.57	35,621	4828
Total				46.96	48,172	5210
yeast_1	4873	528,682	63,682	31.08	17,066	3361
yeast_2	4732	541,194	66,441	30.88	16,921	3325
yeast_3	4691	556,675	66,978	31.98	17,494	3369
Total				31.32	24,131	3627

probability of ions transmitted into the vacuum system of about 10%. This excellent number is because of the fact that the continuous beam generated by the ES source can be utilized and that all the ion guiding elements have been successfully optimized for high transmission.

Resolution and Mass Accuracy—For the impact II several improvements were implemented: symmetrical shielding for better ion focusing; line grids to increase the transmission; low temperature coefficient ceramic spacers to decrease temperature related mass drift and improved axel bearings for precise alignment. Together this led to about 35% increased resolution over the full proteomics mass range. As this improvement is reached solely by better focusing of the ions, mass accuracy and signal to noise are expected to increase accordingly. Improvements to the MCP detector include an increased entrance aperture, higher electron accelerating fields and optimized shielding. Overall, these measures lead to 2-fold faster ion impact transients and 30% higher detection efficiency of the MCP.

In summary, the resolving power of the TOF analyzer is expected to increase by about 70 to 80% by the introduction of the new collision cell, reflectron and detector. To test this experimentally on a standard proteomic sample, we analyzed data from a HeLa digest. Resolution for typical peptides is in excess of 33,000 as illustrated by an example in Fig. 4A. Further increase in the resolution can be obtained by the "Focus mode," which involves real time processing and alignment of successive pulses and increases accuracy of flight time determination, when multiple ions of one species reach the detector at the same time (45). This can be helpful to resolve the isotope distributions of proteins, as shown in Fig. 4B, which features a resolution in excess of 60,000 for an antibody subunit (> 25 kDa). In TOF measurements, resolution tends to increase with *m/z* (Fig. 4C), and reached more than 40,000 for the TuneMix component at *m/z* 1222. This constitutes a 70 to 80% improvement over the previous impact model, the impact HD.

This increase in resolving power should also imply better mass accuracy in proteomics samples, which we tested using the software package MaxQuant, which we adapted to QTOF data as described under "Experimental Proce-

dures." A special feature of MaxQuant is the extraction of individual mass accuracy values (46, 47), which allows to make efficient use of high mass accuracy in the identification of peptides. MaxQuant was originally developed on the basis of data from hybrid Orbitrap instruments. Here we developed MaxQuant further in order to analyze QTOF data and also in this context profit from the high mass accuracy provided by nonlinear mass recalibration algorithms. A special challenge in QTOF data is the drift in the mass scale because of thermal expansion caused by ambient temperature drift. MaxQuant employs a double search strategy, in effect supplying hundreds of reference masses internal to each proteomic sample. This feature efficiently removes any effect of the temperature related mass drift. This "software lock mass" feature makes it unnecessary to use dedicated molecular species for the calibration of spectra (48). With these developments in MaxQuant, we analyzed the peptide mass error distributions over a 90 min gradient run (Fig. 4D). This showed an average absolute mass deviation of around 1.45 ppm, which is excellent for a QTOF instrument.

Impact II Performance for Single Shot Analysis—To investigate the performance of the impact II for shotgun proteomics, we first analyzed a complex peptides mixture derived from a mammalian cell line in the single-run format (Experimental Procedures). We separated 1 μg of peptide digest by on-line HPLC with the standard 90 min gradient employed in our laboratory and performed triplicate analysis. The typical data dependent acquisition scheme in bottom-up proteomics consists of an MS scan followed by N MS/MS fragment scans of the most intense precursors (topN method). It is desirable to choose N such that the total cycle is less than a few seconds. For our measurements we aimed at a duty cycle of around 1.3 s and designed a top17 method, consisting of 200 ms for MS acquisition and a MS/MS integration time adapted to the precursor intensity. We found this to be a good balance between acquiring high S/N in the MS and achieving optimal ion intensities in the MS/MS spectra for the HeLa digest. This method reached more than 7000 MS and 79,700 MS/MS scans. In each run MaxQuant identified on average 35,580 unique peptide sequences, which results in total of 48,172 unique peptide sequences in the triplicate analysis
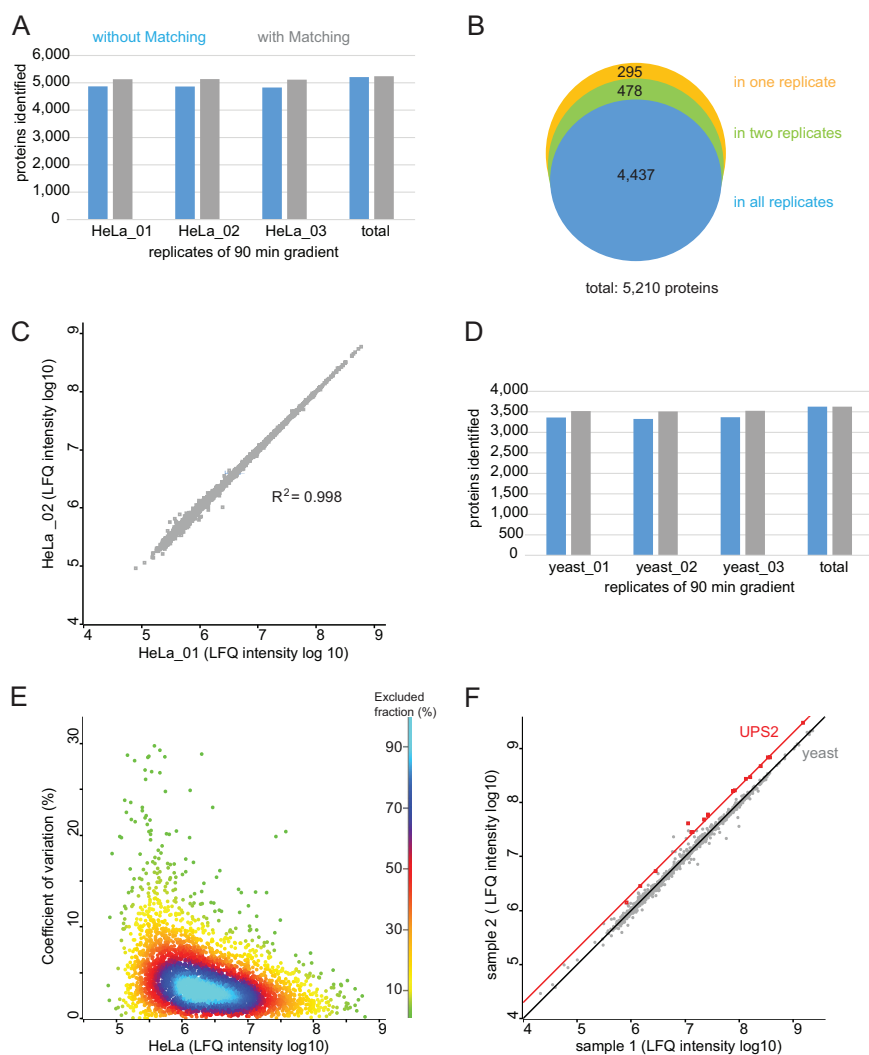
FIG. 5. **Triplicate analysis of HeLa and yeast digest using a 90 min gradient for single shot analysis.** *A*, Protein identification numbers of each replicate of 1 μg HeLa digest and *B*, the overlap of protein identities. *C*, Correlation between the rank ordered label-free quantification values of each identified protein in replicate1 (HeLa_01) and replicate2 (HeLa_02) (log10 LFQ intensities). *D*, Protein identification number of 1 μg yeast digest. *E*, Reproducibility across triplicate analysis of HeLa digest. CVs of all three replicates, representing 99% of the data points. *F*, Accuracy in quantification. Results of spike-in experiment showing the UPS-2 standard (orange) and in a yeast proteome background (gray). Yeast lysate was present in equal amounts in sample 1 and sample 2 and the UPS2 protein standard was present in twice the amount in sample 2.

(Table I). These peptides mapped to an average of 4854 proteins per run, and a total of 5210 proteins of the HeLa proteome with the three 90 min gradients (Fig. 5*A*), indicating that a deep coverage can be achieved using relatively short, single shot analysis. Transferring identification between the runs based on their mass precision and retention time ("match between runs" feature in MaxQuant) led to around 5100 proteins identifications per single run. Comparing protein identities between the triplicate analyses (without "matching between runs"), we observed that more than 90% of proteins were identified in each of them (Fig. 5*B*). This indicates high

reproducibility and a minimal 'missing value' problem. This conclusion is further supported by excellent reproducibility ($R^2 = 0.998$) in the label-free intensities determined in pairwise comparison between runs (MaxQuant LFQ values (39, 49)) (Fig. 5*C*). We also determined the number of proteins identified in a single shot run of 1 μg yeast digest. On average, we identified 3352 proteins per single, 90 min gradient, and a total of 3627 proteins when combining the three single shot measurements (Fig. 5*D*).

Reproducibility and Accuracy of Quantification—To evaluate the reproducibility of the method for label-free quantifica-
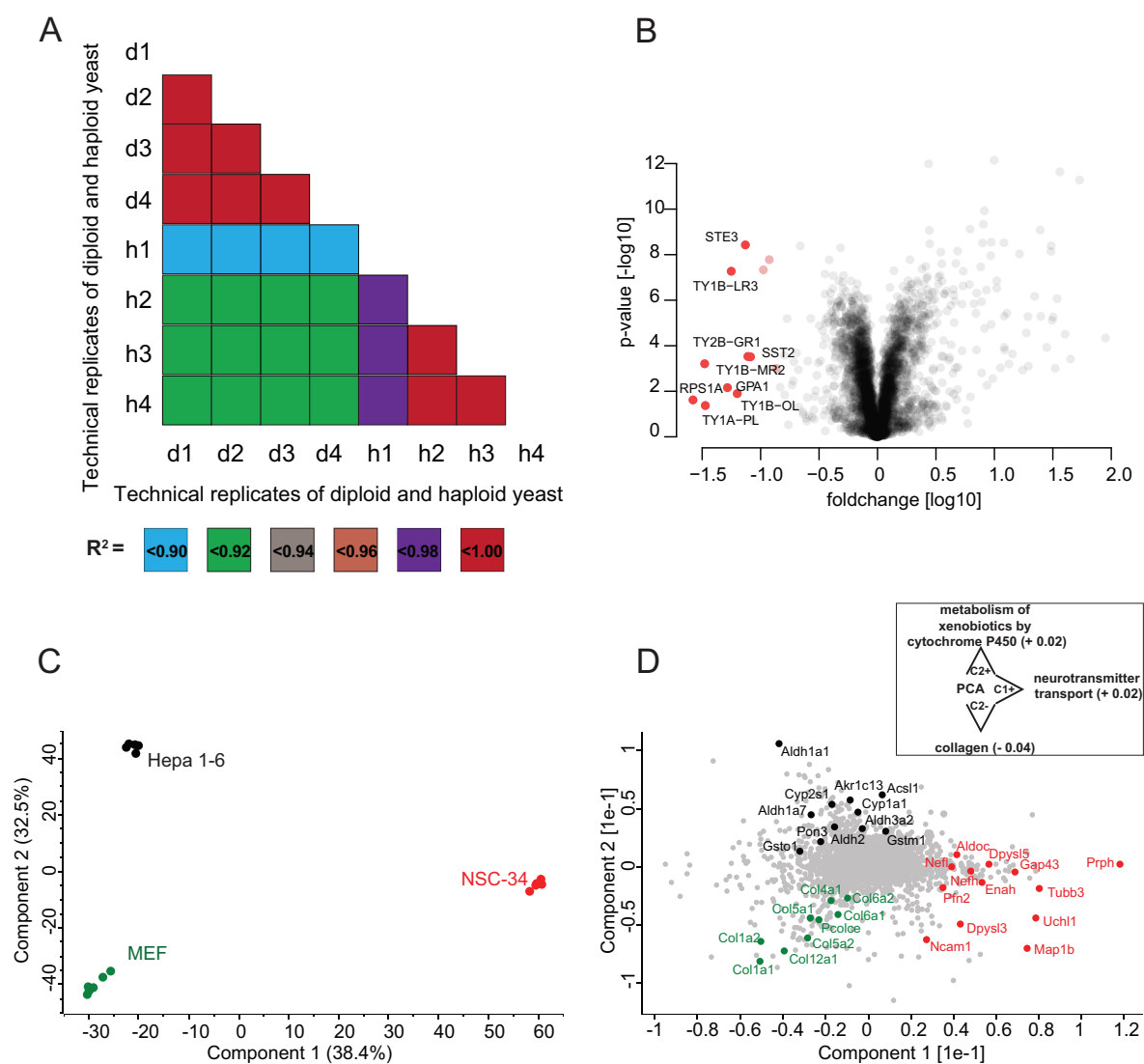
Fig. 6. *A*, **Correlation of single shot LC-MS/MS measurements of haploid (h) and diploid (d) yeast samples.** All technical replicate correlation values were > 0.98. *B*, Quantitative differences between the haploid and diploid yeast proteome. Proteins marked in red are significantly more abundant in haploid cells. *C*, Principal component analysis (PCA) of LFQ protein expression levels in mouse hepatoma (Hepa 1–6), mouse embryonic fibroblasts (MEF) and motoneuronal cell lines (NSC-34). *D*, Loading of the first two principal components as obtained from the PCA plotted in *C*. The inset indicates significantly enriched annotations along the PC axes (FDR < 0.05). The respective enrichment scores are specified in parentheses.

tion, we determined the coefficients of variation (CV) of the label-free intensities, determined in pair-wise comparison between three technical HeLa replicates (see above). For more than 90% of the quantified proteins the CV was smaller than 10% (Fig. 5*E*). For the lowest intensity quantile the median was 0.05 and for the highest one 0.03 (Fig. 5*E*, supplemental Fig. S3). This shows reproducible quantification over four orders of magnitude.

Accurate quantification of differently expressed proteins remains challenging over a wide concentration range and

benefits from a very stable analytical platform. QTOF instrumentation has been used widely for label-free quantification, which in principle allows convenient analysis and comparison of an arbitrary number of samples. For evaluation of the label-free quantitative capabilities of the impact II platform in complex mixtures we wished to use a reference sample set with known ratios for a small subset of proteins. We spiked the Universal Protein Standard 2 (UPS-2), consisting of 48 proteins covering a dynamic range of five orders of magnitude, in two different concentrations into the yeast proteome.

This generated two samples, in which the yeast peptides should be 1:1 whereas ratios for UPS-2 peptides should be at 1:2. To increase the number of identified peptides we then used the equimolar UPS-1, which we also spiked in a yeast background. This allowed transfer of peptide identifications to unsequenced peptides in the UPS-2 in yeast runs via the 'match between runs' algorithm in MaxQuant. We identified all of the 48 human UPS proteins in the sample containing yeast with the equimolar UPS-1 standard. Of these proteins, 23 were identified in each one of the 12 single shot yeast measurements with the dynamic UPS-2 standard either directly via MS/MS or via match between runs and 18 of these proteins showed Welch-significance (Fig. 5F). We identified and quantified UPS-2 proteins over more than three orders of magnitude in these relatively fast measurements (90 min gradients). The UPS-2 proteins have an average fold change of 0.49 ($\pm$ 0.06), which is close to the theoretical ratio of 0.5.

Quantification of Changes in the Yeast Proteome—To test the workflow in a systems biology context, we analyzed proteome changes of diploid and haploid (Matα cell) *S. cerevisiae*. We analyzed 2 μg of yeast digest from haploid and diploid cells in technical quadruplicates with our standard 90 min gradient. This identified 3769 proteins using "match between runs." For statistical analysis, we only considered LFQ intensities that were detected in at least three replicates of the haploid or diploid groups. After filtering, 3222 proteins remained for further analysis (Experimental Procedures). Remaining missing values were imputed from a normal distribution. The technical replicates correlated much more with each other than they correlated to the other genotype (R^2 greater 0.98 *versus* R^2 about 0.92; Fig. 6A). As in our previous large-scale analysis on SILAC labeled haploid and diploid yeast (50), we found transposons more abundant in haploid cells than in diploid cells (Fig. 6B). Ste3, the pheromone a factor receptor, was specific to haploid yeast, as expected from its mating status. Also absent in diploid but present in haploid cells were Sst2, a GTPase-activating protein for Gpa1 (51), which, consistently, showed higher expression in haploid cells. Conversely, Sps100, which is a sporulation-specific wall maturation protein turned out to be specific for diploid cells. Doing such systems-wide comparisons by traditional methods would have required thousands of individual Western blots. Even compared with our previous large-scale study performed by quantitative MS (50), we here used less than 1% of yeast input material and measurement time. This illustrates the rapidity by which MS-based proteomics is becoming a viable method for answering biological questions.

Global Proteomic Comparison of Different Cell Lines—As a second example of typical proteomics experiments, we applied the QTOF-based workflow to the characterization of common cellular disease model systems. For this purpose, we compared the proteomes of spinal cord neuron-neuroblastoma (NSC-34), mouse hepatoma (Hepa 1–6) and mouse embryonic fibroblast (MEF) cell lines in a quantitative manner.

All cell lysates were analyzed in single shots using 90 min gradients and subsequently quantified using the MaxQuant label-free quantification algorithm. The observed LFQ intensities were highly reproducible between biological and technical replicates as indicated by Pearson correlations coefficients > 0.97 (supplemental Fig. S4).

After stringent filtering (Experimental Procedures) we performed a principal component analysis (PCA) to evaluate the similarities and dissimilarities of the cell lines on a global scale. Replicates from a single cell line clustered very tightly in the PCA space and the first two principal components accounted for 71% of the cumulative variance within our data set (Fig. 6C). Interestingly, the motoneuronal cell line NSC-34 was clearly separated from Hepa 1–6 and MEF cells in the first principle component, whereas the variance between both non-neuronal cell lines was described by the second, orthogonal, principal component. In the latter component, NSC-34 is positioned half-way between Hepa 1–6 and MEF.

To assess individual proteins that are the main drivers for the separation between the three cell lines, we plotted the loadings of the first two principal components (Fig. 6D). Hepa 1–6 cells were characterized by gene products involved in regulatory and metabolic processes. As we had found before (52), proteins related to the glycolysis pathway, such as the aldehyde dehydrogenase (Aldh) family, were highly represented in Hepa 1–6. Not surprisingly, proteins driving the separation of MEF were predominantly linked to collagen synthesis, such as the COL gene family and the precollagen C-endopeptidase enhancer 1 (Pcolce). Proteins that differentiate along principal component 1 encompassed various components of the cytoskeleton axons, including Prph and the heavy and light chains of neurofilaments (Nefh and Nefl). In addition, gene products involved in axon guidance (Enah, Dypsl5 and Tubb3) and neuron projection (Uchl1, Gap43) were highly distinctive for NSC-34. Proteins separating NSC-34 from the other cell lines were significantly enriched for neurotransmitter transport while we observed enrichment for collagens and xenobiotic processes for the MEF and Hepa 1–6 cells on component 2, respectively, further showing how proteomics can highlight biological function. This is even more remarkable, given that we previously found that the proteomes of motoneuronal cell lines, including NSC-34, lack distinctive neuronal characteristics, as several key actors in axon growth and guidance were either depleted or low abundant (28). As a result, we had placed motoneuronal cell lines only halfway between *in vivo* motoneurons and non-neuronal controls. Nevertheless, our label-free QTOF-based workflow is very well suited to differentiate subtle alterations in biological systems in a short time of analysis.

Impact II Performance for Deep Proteome Analysis of a Cell Line—To evaluate the impact II for deep proteome coverage we performed high pH reverse-phase pre-fractionation with fraction concatenation as described in ref (31). We loaded 175 μg of a HeLa peptide mixture, collected 72 fractions and
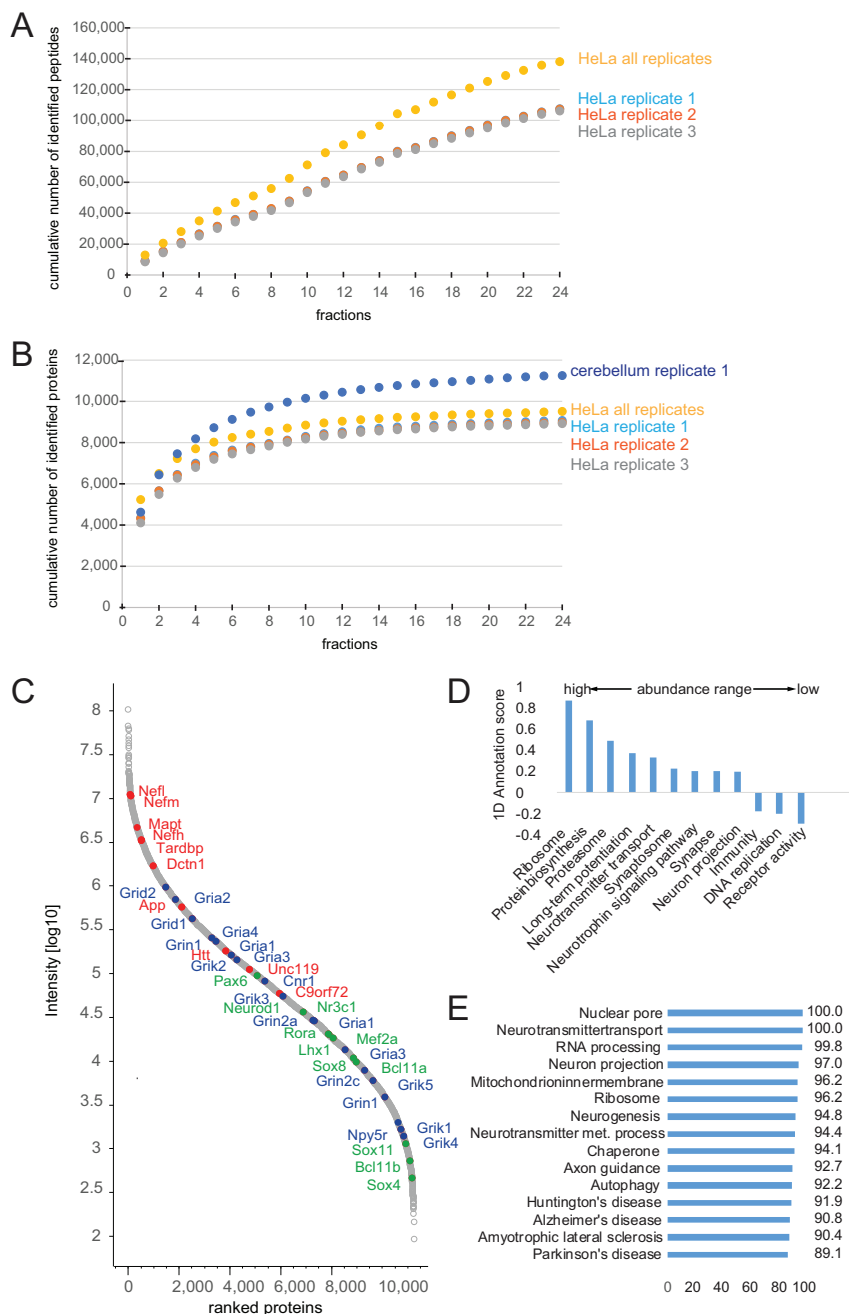
FIG. 7. **Triplicate cell line and singlet tissue analysis in 24 high-pH fractions.** *A*, Cumulative number of identified peptides of triplicate analysis of HeLa. *B*, Cumulative number of identified proteins of HeLa replicates and of the singlet analysis of mouse cerebellum (48 h total measurement time). *C*, Rank ordered intensities of each identified protein (log10 intensities) of cerebellum. Proteins involved in neurodegenerative disorders are marked in red; cerebellum specific expressed proteins are marked in blue and a selection of receptors in green. *D*, Protein abundance distribution and *E*, proteome coverage (expressed as a percentage) of different neuronal, metabolic and disease relevant processes in cerebellum.

combined them into 24 fractions (Experimental Procedures). These fractions were analyzed in technical triplicates using a standard 90 min LC-MS/MS gradient (total measurement time of 48 h per replicate). Instead of the simple Top17 method described above, we here used a so-called "dynamic method" with a fixed cycle time and a MS/MS integration time adapted to the precursor intensity. We found that this method helped to generate high quality MS/MS spectra also for low abundant peptides.

Figure 7A depicts the cumulative number of unique peptides identified as a function of the number of fractions analyzed. The increase is nearly linear, indicating a small overlap of peptide identification between fractions and the good orthogonal separation power as also observed by others using high-pH fractionation (49). The figure also indicates that the reproducibility was very high between the technical replicates. On average we identified 107,038 unique peptides, mapping to around 8995 proteins (protein FDR 1%; Fig. 7B). In total we identified 138,086 unique peptides, resulting in 9515 different protein groups. This is to our knowledge the deepest proteome coverage of a human cell line measured with a QTOF instrument.

Impact II Performance for Ultra-Deep Proteome Analysis of Tissue—Tissues are more challenging to analyze by proteomics than cell lines, because they are made up of different cell types, the extracellular matrix and other structural and connective elements. Among the different tissues, the brain is thought to be the most complex one. To evaluate the Impact II in this context, we homogenized the cerebellum of a single mouse, digested it using our standard workflow and separated part of the resulting tryptic peptides using high pH fractionation (Experimental Procedures). In total, we identified 11,257 proteins from a single analysis of 24 fractions (2 days of measurement time) (Fig. 7B, supplemental Table S1). To our knowledge, this is the deepest proteome measurement of any tissue reported by TOF instrumentation so far. Protein abundance, as indicated by the summed and normalized peptide signal varied by more than five orders of magnitudes (Fig. 7C). We identified many high to medium abundant proteins involved in neurodegenerative disorders (Fig. 7C marked in red). The transcription factors and DNA binders Pax6, Lhx1/5, Otx1/2, and Neurod1 (Fig. 7C, marked in blue) are examples of proteins that have been reported to be specifically expressed in cerebellum (53). Like various neuronal receptors, transcription factors populate the medium to low abundance range (Fig. 7C, marked in green). The distribution of molecular functions throughout the abundance range is similar to that observed in previous studies (Fig. 7D, supplemental Table S2) (54, 55). The very high depth of this cerebellum proteome is shown by almost complete coverage for neuronal, general metabolic and disease relevant processes (Fig. 7E, supplemental Table S2). Rapid estimation of protein abundances in the brain regions of a single mouse could be useful for studying tissue characteristics and disease specific alterations. For instance, knowledge about changes in the complete pro-

teome would provide an additional layer of information on the pathological processes in neurodegeneration.

CONCLUSION

Here we have described the construction and performance of a state of the art QTOF instrument, the impact II. We documented significant improvements in the ion path, collision cell performance and, in particular, in the performance of the reflectron and detector. The latter allow a mass resolution and mass accuracy that is compatible with the high demands of shotgun proteomics of complex mixtures. For the first time, we measured and modeled the ion transmission from capillary to the detector, which revealed an excellent efficiency of about 10%. The new features of the impact II allow deep characterization in single shots, where we identified more than 5200 proteins in HeLa cells and 3600 proteins in yeast. Using off-line high pH reversed-phase fractionation we identified more than 9500 proteins in HeLa cells and 11,250 proteins in a single cerebellum tissue analysis. These are extremely high numbers for any platform and additional method developments should further improve these results. We also document excellent quantitative reproducibility and accuracy in a label-free format. In concordance with others (26, 27), we conclude that the improvements in QTOF technologies in recent years now clearly enable demanding, in-depth analysis of very complex proteomes.

REFERENCES

1. Karas, M., and Hillenkamp, F. (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* **60,** 2299–2301
2. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246,** 64–71
3. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422,** 198–207
4. Cravatt, B. F., Simon, G. M., and Yates, J. R., 3rd (2007) The biological impact of mass-spectrometry-based proteomics. *Nature* **450,** 991–1000
5. Altelaar, A. F., Munoz, J., and Heck, A. J. (2013) Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat. Rev. Genet.* **14,** 35–48
6. Richards, A. L., Merrill, A. E., and Coon, J. J. (2014) Proteome sequencing goes deep. *Curr. Opin. Chem. Biol.* **24C,** 11–17
7. Domon, B., and Aebersold, R. (2006) Challenges and opportunities in

proteomics data analysis. *Mol. Cell. Proteomics* **5,** 1921–1926

8. Mann, M., and Kelleher, N. L. (2008) Precision proteomics: the case for high-resolution and high mass-accuracy. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 18132–18138

9. Makarov, A. (2000) Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal. Chem.* **72,** 1156–1162

10. Zubarev, R. A., and Makarov, A. (2013) Orbitrap mass spectrometry. *Anal. Chem.* **85,** 5288–5296

11. Michalski, A., Damoc, E., Hauschild, J. P., Lange, O., Wieghaus, A., Makarov, A., Nagaraj, N., Cox, J., Mann, M., and Horning, S. (2011) Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol. Cell. Proteomics* **10,** M111 011015

12. Scheltema, R. A., Hauschild, J. P., Lange, O., Hornburg, D., Denisov, E., Damoc, E., Kuehn, A., Makarov, A., and Mann, M. (2014) The Q exactive HF, a benchtop mass spectrometer with a pre-filter, high-performance quadrupole and an ultra-high-field Orbitrap analyzer. *Mol. Cell. Proteomics* **13,** 3698–3708

13. Kelstrup, C. D., Jersie-Christensen, R. R., Batth, T. S., Arrey, T. N., Kuehn, A., Kellmann, M., and Olsen, J. V. (2014) Rapid and deep proteomes by faster sequencing on a benchtop quadrupole ultra-high-field Orbitrap mass spectrometer. *J. Proteome Res.* **13(12),** 6187–6195

14. Zhou, H., Ning, Z., Wang, F., Seebun, D., and Figeys, D. (2011) Proteomic reactors and their applications in biology. *FEBS J.* **278,** 3796–3806

15. Kocher, T., Swart, R., and Mechtler, K. (2011) Ultra-high-pressure RPLC hyphenated to an LTQ-Orbitrap Velos reveals a linear relation between peak capacity and number of identified peptides. *Anal. Chem.* **83,** 2699–2704

16. Cox, J., and Mann, M. (2011) Quantitative, high-resolution proteomics for data-driven systems biology. *Annu. Rev. Biochem.* **80,** 273–299

17. Nagaraj, N., Kulak, N. A., Cox, J., Neuhauser, N., Mayr, K., Hoerning, O., Vorm, O., and Mann, M. (2012) System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. *Mol. Cell. Proteomics* **11,** M111 013722

18. Hebert, A. S., Richards, A. L., Bailey, D. J., Ulbrich, A., Coughlin, E. E., Westphall, M. S., and Coon, J. J. (2014) The one hour yeast proteome. *Mol. Cell. Proteomics* **13,** 339–347

19. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N., and Mann, M. (2014) Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **11,** 319–324

20. Michalski, A., Cox, J., and Mann, M. (2011) More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *J. Proteome Res.* **10,** 1785–1793

21. Savitski, M. M., Nielsen, M. L., and Zubarev, R. A. (2006) ModifiComb, a new proteomic tool for mapping substoichiometric post-translational modifications, finding novel types of modifications, and fingerprinting complex protein mixtures. *Mol. Cell. Proteomics* **5,** 935–948

22. Morris, H. R., Paxton, T., Dell, A., Langhorne, J., Berg, M., Bordoli, R. S., Hoyes, J., and Bateman, R. H. (1996) High sensitivity collisionally-activated decomposition tandem mass spectrometry on a novel quadrupole/orthogonal-acceleration time-of-flight mass spectrometer. *Rapid Commun. Mass Spectr.* **10,** 889–896

23. Silva, J. C., Denny, R., Dorschel, C., Gorenstein, M. V., Li, G. Z., Richardson, K., Wall, D., and Geromanos, S. J. (2006) Simultaneous qualitative and quantitative analysis of the Escherichia coli proteome: a sweet tale. *Mol. Cell. Proteomics* **5,** 589–607

24. Silva, J. C., Gorenstein, M. V., Li, G. Z., Vissers, J. P., and Geromanos, S. J. (2006) Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol. Cell. Proteomics* **5,** 144–156

25. Gillet, L. C., Navarro, P., Tate, S., Rost, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012) Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11,** O111 016717

26. Distler, U., Kuharev, J., Navarro, P., Levin, Y., Schild, H., and Tenzer, S. (2014) Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics. *Nat. Methods* **11,** 167–170

27. Helm, D., Vissers, J. P., Hughes, C. J., Hahne, H., Ruprecht, B., Pachl, F., Grzyb, A., Richardson, K., Wildgoose, J., Maier, S. K., Marx, H., Wilhelm, M., Becher, I., Lemeer, S., Bantscheff, M., Langridge, J. I., and Kuster, B. (2014) Ion mobility tandem mass spectrometry enhances performance of bottom-up proteomics. *Mol. Cell. Proteomics* **13,** 3709–3715

28. Hornburg, D., Drepper, C., Butter, F., Meissner, F., Sendtner, M., and Mann, M. (2014) Deep proteomic evaluation of primary and cell line motoneuron disease models delineates major differences in neuronal characteristics. *Mol. Cell. Proteomics* **13,** 3410–3420

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

30. Wang, H., Qian, W. J., Mottaz, H. M., Clauss, T. R., Anderson, D. J., Moore, R. J., Camp, D. G., 2nd, Khan, A. H., Sforza, D. M., Pallavicini, M., Smith, D. J., and Smith, R. D. (2005) Development and evaluation of a micro- and nanoscale proteomic sample preparation method. *J. Proteome Res.* **4,** 2397–2403

31. Dwivedi, R. C., Spicer, V., Harder, M., Antonovici, M., Ens, W., Standing, K. G., Wilkins, J. A., and Krokhin, O. V. (2008) Practical implementation of 2D HPLC scheme with accurate peptide retention prediction in both dimensions for high-throughput bottom-up proteomics. *Anal. Chem.* **80,** 7036–7042

32. Whitehouse, C. M., Dreyer, R. N., Yamashita, M., and Fenn, J. B. (1985) Electrospray interface for liquid chromatographs and mass spectrometers. *Anal. Chem.* **57,** 675–679

33. Franzen, J. (1998) Method and device for transport of ions in gas through a capillary. Google Patents

34. Ishihama, Y., Rappsilber, J., Andersen, J. S., and Mann, M. (2002) Microcolumns with self-assembled particle frits for proteomics. *J. Chromatogr. A* **979,** 233–239

35. Ayoub, D., Jabs, W., Resemann, A., Evers, W., Evans, C., Main, L., Baessmann, C., Wagner-Rousset, E., Suckau, D., and Beck, A. (2013) Correct primary structure assessment and extensive glyco-profiling of cetuximab by a combination of intact, middle-up, middle-down, and bottom-up ESI and MALDI mass spectrometry techniques. *mAbs* **5,** 699–710

36. Senko, M. W., Beu, S. C., and McLaffertycor, F. W. (1995) Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. *J. Am. Soc. Mass Spectr.* **6,** 229–233

37. Tsybin, Y. O., Fornelli, L., Stoermer, C., Luebeck, M., Parra, J., Nallet, S., Wurm, F. M., and Hartmer, R. (2011) Structural analysis of intact monoclonal antibodies by electron transfer dissociation mass spectrometry. *Anal. Chem.* **83,** 8919–8927

38. Cox, J., Neuhauser, N., Michalski, A., Scheltema, 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

39. Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., and Mann, M. (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13,** 2513–2526

40. R Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

41. 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,** S12

42. Shaffer, S. A., Prior, D. C., Anderson, G. A., Udseth, H. R., and Smith, R. D. (1998) An ion funnel interface for improved ion focusing and sensitivity using electrospray ionization mass spectrometry. *Anal. Chem.* **70,** 4111–4119

43. Gerlich, D. (1992) *Inhomogeneous RF fields a versatile tool for the study of processes with slow ions.*, John Wiley & Sons, Inc, Hoboken, NJ, U.S.A.

44. Steen, H., Kuster, B., Fernandez, M., Pandey, A., and Mann, M. (2001) Detection of tyrosine phosphorylated peptides by precursor ion scanning quadrupole TOF mass spectrometry in positive ion mode. *Anal. Chem.* **73,** 1440–1448

45. Räther, O. (2005) High resolution detection for time-of-flight mass spectrometers. US6870156 B2 patent

46. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnol.* **26,** 1367–1372

47. Cox, J., Hubner, N. C., and Mann, M. (2008) How much peptide sequence information is contained in ion trap tandem mass spectra? *J. Am. Soc.*

Mass Spectr. **19,** 1813–1820

48. Cox, J., Michalski, A., and Mann, M. (2011) Software lock mass by two-dimensional minimization of peptide mass errors. *J. Am. Soc. Mass Spectr.* **22,** 1373–1380

49. Wang, Y., Yang, F., Gritsenko, M. A., Wang, Y., Clauss, T., Liu, T., Shen, Y., Monroe, M. E., Lopez-Ferrer, D., Reno, T., Moore, R. J., Klemke, R. L., Camp, D. G., 2nd, and Smith, R. D. (2011) Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *Proteomics* **11,** 2019–2026

50. de Godoy, L. M., Olsen, J. V., Cox, J., Nielsen, M. L., Hubner, N. C., Frohlich, F., Walther, T. C., and Mann, M. (2008) Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **455,** 1251–1254

51. Apanovitch, D. M., Slep, K. C., Sigler, P. B., and Dohlman, H. G. (1998) Sst2 is a GTPase-activating protein for Gpa1: purification and characterization of a cognate RGS-Galpha protein pair in yeast. *Biochemistry* **37,** 4815–4822

52. Pan, C., Kumar, C., Bohl, S., Klingmueller, U., and Mann, M. (2009) Comparative proteomic phenotyping of cell lines and primary cells to assess preservation of cell type-specific functions. *Mol. Cell. Proteomics* **8,** 443–450

53. Suzuki, H., Okunishi, R., Hashizume, W., Katayama, S., Ninomiya, N., Osato, N., Sato, K., Nakamura, M., Iida, J., Kanamori, M., and Hayashizaki, Y. (2004) Identification of region-specific transcription factor genes in the adult mouse brain by medium-scale real-time RT-PCR. *FEBS Letters* **573,** 214–218

54. Geiger, T., Wehner, A., Schaab, C., Cox, J., and Mann, M. (2012) Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol. Cell. Proteomics* **11,** M111 014050

55. Schwanhausser, 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

Article 6

The Q Exactive HF, a benchtop mass spectrometer with a pre-filter, high-performance quadrupole and an ultra-high-field Orbitrap analyzer

A major limiting factor in proteomics is the performance of the mass spectrometer. Because in proteomics there is no amplification step like in oligonucleotide-based sequencing technologies, the sensitivity, dynamic range, resolution and speed of the mass spectrometer has a very large impact on the robustness, depth and accuracy of the proteome analysis. The introduction of the benchtop quadrupole Orbitrap mass spectrometer (Q Exactive) has made a compact and powerful instrument available to the biological community for quantitative proteomics. We optimized and benchmarked the latest type of hybrid Orbitrap mass spectrometer, and determined the parameters for peak performance. We used a Design of Experiment (DoE) approach to determine these optimal parameters in a robust and unbiased manner. In contrast to the conventional strategy of 'changing one parameter at a time', DoE allows to determine higher order dependencies of parameters systematically with a minimal set of experiments. With the optimized set of parameters, we were able to identify 5,000 proteins within 90 min mass spectrometric analysis of mammalian cell lysate.

Molecular & Cellular Proteomics (2014) [5]

Scheltema RA, Hauschild JP, Lange O, **Hornburg D**, Denisov E, Damoc E, Kuehn A, Makarov A, Mann M.

Contribution: In this project, Thermo Fisher and the group of Matthias Mann (MPI of Biochemistry) presented, benchmarked and optimized the latest generation of hybrid quadrupole Orbitrap MS for proteomics research. To assess the complexity and dependencies of parameters that affect proteome analyses in the Q Exactive HF, I designed experiments to quantify and model the effects of various instrument parameter on the machine performance. I conducted the measurements, data analysis and interpretation and employed a central composite faced design (CCF, DoE) to model a response surface for parameter settings. This approach provided confident information on the parameters dependencies and their optimal settings.

The Q Exactive HF, a Benchtop Mass Spectrometer with a Pre-filter, High-performance Quadrupole and an Ultra-high-field Orbitrap Analyzer*⑤

Richard Alexander Scheltema‡, Jan-Peter Hauschild§, Oliver Lange§, Daniel Hornburg‡, Eduard Denisov§, Eugen Damoc§, Andreas Kuehn§, Alexander Makarov§, and Matthias Mann‡¶

The quadrupole Orbitrap mass spectrometer (Q Exactive) made a powerful proteomics instrument available in a benchtop format. It significantly boosted the number of proteins analyzable per hour and has now evolved into a proteomics analysis workhorse for many laboratories. Here we describe the Q Exactive Plus and Q Exactive HF mass spectrometers, which feature several innovations in comparison to the original Q Exactive instrument. A low-resolution pre-filter has been implemented within the injection flatapole, preventing unwanted ions from entering deep into the system, and thereby increasing its robustness. A new segmented quadrupole, with higher fidelity of isolation efficiency over a wide range of isolation windows, provides an almost 2-fold improvement of transmission at narrow isolation widths. Additionally, the Q Exactive HF has a compact Orbitrap analyzer, leading to higher field strength and almost doubling the resolution at the same transient times. With its very fast isolation and fragmentation capabilities, the instrument achieves overall cycle times of 1 s for a top 15 to 20 higher energy collisional dissociation method. We demonstrate the identification of 5000 proteins in standard 90-min gradients of tryptic digests of mammalian cell lysate, an increase of over 40% for detected peptides and over 20% for detected proteins. Additionally, we tested the instrument on peptide phosphorylation enriched samples, for which an improvement of up to 60% class I sites was observed. *Molecular & Cellular Proteomics 13: 10.1074/mcp.M114.043489, 3698–3708, 2014.*

From the ‡Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Am Klopferspitz 18, D-82152 Martinsried, Germany; §Thermo Fisher Scientific (Bremen) GmbH, Hanna-Kunath-Strasse 11, 28199 Bremen, Germany

Mass spectrometry (MS)-based[1] proteomics aims at the comprehensive analysis of proteins present in a biological sample (1), and the field has expanded in many surprising directions (2). Application of the developed techniques has revealed novel insights into fundamental biology, as well as produced analysis techniques with implications for clinical applications. A major hurdle, however, is the complexity of the systems under scrutiny, as it has been shown that human cell lines, for instance, express at least 10,000 genes that are detectable as proteins (3–5). If we further consider all the peptides produced in bottom-up proteomics experiments, this hurdle is compounded, as ideally many hundreds of thousands of analytes should be characterized in order for the proteins giving rise to them to be fully reconstructed (6). In principle, issues of sample complexity and dynamic range could be addressed by a very high degree of up-front fractionation. However, this strategy faces diminishing returns and leads to unacceptably long analysis times for most purposes. Given the fact that even with optimal chromatographic resolution many peptides with abundance differences of many orders of magnitude elute within the same time frame, there remains a need to improve the mass spectrometric detection in terms of speed, resolution, and sensitivity.

Nanoscale liquid chromatography coupled online to mass spectrometry is the current technique of choice for the analysis of complex peptide mixtures. In a top-*N* shotgun strategy, a full scan, providing a complete overview of isotope patterns resulting from ionized peptides, is followed by *N* fragmentation scans performed on the most abundant not-yet-sequenced isotope patterns currently visible in the full scan. During fragmentation, the goal is to cleanly isolate the intended precursor peptide ion, which today is generally done either by a linear ion trap or by a quadrupole mass filter. Fragment ions are then mass measured by an Orbitrap mass analyzer, a time-of-flight analyzer, or, less often, ion cyclotron
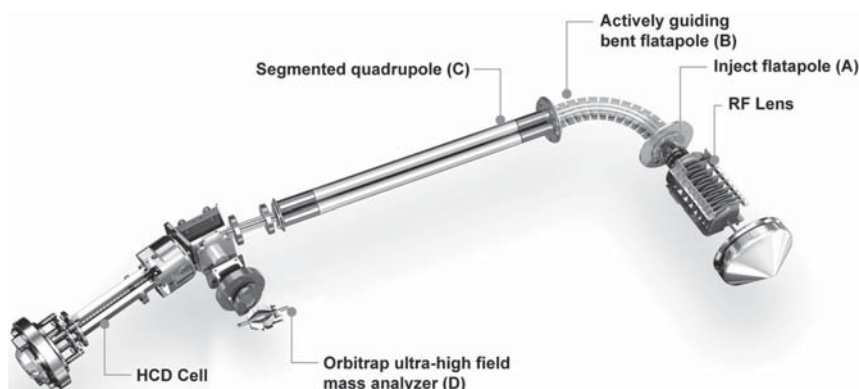

FIG. 1. **Construction details of the Q Exactive HF.** This instrument is based on the Q Exactive series and improves on it with a mass selection pre-filter implemented at the injection flatapole (*A*), an actively guiding bent flatapole (*B*), and a segmented quadrupole (*C*). This combination prevents contamination from traveling far into the instrument and improves the ion transmission almost 2-fold. The ultra-high-field Orbitrap analyzer (*D*) is optional.

resonance–Fourier transform or linear or three-dimensional ion traps.

Apart from the MS instrumentation, recent developments in the proteomics workflow include a move toward automated online quality control systems (7, 8) and single-run analyses (9), which require very high-performance peptide chromatography (10, 11).

The Orbitrap mass analyzer was introduced commercially almost 10 years ago, and hybrid instruments based on this tool have become very popular in proteomics (12). They consist of an upfront mass spectrometer coupled to a so-called C-trap, which stores and compresses the ion population (generally up to one million charges) prior to injection into the Orbitrap analyzer. Up to the Orbitrap Velos and Elite members of this family of instruments, the precursor selection (and usually the fragmentation) occurred in the linear ion trap (13), but a few years ago an instrument based on a quadrupole front end—the Q Exactive mass spectrometer—was developed (14). Compared with the linear ion trap, quadrupole mass filters have the advantage of being capable of nearly instantaneously selecting a small mass region by modulating the RF field, allowing only a select set of ions to have stable trajectories when passing through the rod assembly. As we described previously, this near-instantaneous mass selection capability and the C-trap's capability of storing ions enable multiplexing of different ion populations (*e.g.* fragment ions of two or more distinct precursor ions) prior to analysis in the Orbitrap mass analyzer (14). However, that instrument did not use the highest efficiency quadrupole technology (15), and it did not use the compact high-field Orbitrap analyzer that had been introduced in the Orbitrap Elite instrument (16).

Here we describe the advances incorporated into the Q Exactive Plus and the Q Exactive HF instruments. These include improved robustness effected by a low-resolution filter upstream of the quadrupole, a segmented quadrupole, and, in the case of the Q Exactive HF instrument, an ultra-high-field mass Orbitrap analyzer, doubling resolution or acquisition speed. We describe these capabilities in the context of single-shot complex mixture analysis of peptides and phosphopeptides.

EXPERIMENTAL PROCEDURES

Construction of the Q Exactive Plus—The instrument is based on the previous-generation Q Exactive mass spectrometer (14) with the same general design elements. Briefly, these consist of an atmospheric pressure ion source, a stacked-ring ion guide (S-lens) and an injection flatapole in the source region, a bent flatapole containing a large bore at the bend ejecting solvent droplets and other neutral species to prevent them from entering further into the instrument, a segmented quadrupole mass filter, a C-trap, a higher energy collisional dissociation cell, and an Orbitrap mass analyzer (Fig. 1). In contrast to the previous generation, the injection flatapole was equipped with ion selection capabilities, providing a low-resolution selection mechanism for the removal of undesirable ions from the ion beam before higher resolution selection in the quadrupole mass filter. The new, segmented quadrupole mass filter has improved ion transmission and mass selection characteristics (see below). Lastly, the instrument can optionally be equipped with a compact Orbitrap analyzer (Q Exactive HF), leading to a higher field and consequently higher frequency of ion motion. This in turn leads to almost twice the resolution in the same scan time relative to the previous cell (16). Alternatively, this analyzer can be used to achieve twice the scan speed while remaining at nearly the same resolution, a strategy we utilize here.

The covered mass range in the Orbitrap analyzer is *m/z* 50–6000. Precursor mass selection by the quadrupole is possible up to *m/z* 2500, and isolation windows can be set between 0.4 and 5600 Th. The instrument software automatically adjusts the required ion injection time to compensate for loss of transmission when reducing the isolation width. Acquisition speed with the classic Orbitrap analyzer remains the same as previously reported, whereas for the high-field Orbitrap it ranges from 27 Hz for resolving power 15,000 specified at *m/z* 200 (corresponding to 10,000 at *m/z* 400) to 1.5 Hz for resolving power 240,000 at *m/z* 200 (corresponding to 170,000 at *m/z* 400). The vacuum in the Orbitrap compartment can be electronically adjusted ~5-fold, enabling high-resolution analysis of most analytes including large peptides and small proteins.

Preparation of HeLa Lysates—HeLa cells (ATCC, S3 subclone) were cultured in DMEM containing 10% fetal bovine serum, 20 mM glutamine, and 1% penicillin-streptomycin. Cells were collected via centrifugation at 200g for 10 min, washed once with cold PBS, and centrifuged again. Supernatant was carefully discarded, and cell pellets were shock frozen in liquid nitrogen and stored at −80 °C. Aliquots of ~3e7 cells were re-suspended in 1 ml of water; then 1 ul of trifluoroethanol was added, and samples underwent incubation for 10 min on ice, a 2-min sonication at duty cycle 30% and output control 3 (Branson, Danbury; sonifier model 250), and 1 min of vortexing. After 20 min of incubation at 56 °C, 25 μl of 200 mM DTT was added to reduce proteins, and samples were incubated at 90 °C for 15 min. Alkylation was then performed by adding 100 μl of 200 mM iodoacetamide and incubating for 60 min at room temperature in the dark. The sample was diluted using 8 ml of 50 mM NH$_4$HCO$_3$ to reduce the final trifluoroethanol concentration to 10% (v/v), after which the sample was digested for 1 h at 37 °C by LysC at an enzyme:protein ratio of 1:100 and then overnight at 37 °C after the addition of trypsin at a ratio of 1:100. Digests were then diluted 1:4 with 0.1% formic acid (v/v) and purified with Sep-Pak tC$_{18}$ cartridges according to the manufacturer's instructions. The peptide concentration was determined using a NanoDrop spectrophotometer (Thermo Scientific).

Phosphorylation Enrichment of Non-stimulated HeLa Cells—Peptides were collected from a population of 1e8 unstimulated HeLa cells prepared using the filter-aided sample preparation method (17). In brief, cell pellets were solubilized in 4% SDS, 100 mM Tris/HCl, pH 7.6, 0.1 M DTT; incubated at 95 °C for 5 min; and sonicated at duty cycle 30% and output control 3 (Branson Ultrasonics). The protein concentration was determined from tryptophan fluorescence emission at 350 nm using an excitation wavelength of 295 nm. A total of 30 mg of protein extract was then split on top of five 30,000 molecular weight cutoff centrifugal filters (20 mg per filter), spun down, and washed twice with 7 ml of 8 M urea, 100 mM Tris/HCl, pH 8.5. Alkylation was performed with 50 mM iodoacetamide for 30 min at room temperature in the dark in the same buffer. After two further washes with 7 ml of 8 M urea, Tris/HCl, pH 8.5, in 0.1 M and three with 7 ml of NH$_4$HCO$_3$, digestion was performed by adding LysC at an enzyme:protein ratio of 1:50 and incubating overnight at 30 °C. The digested peptides were eluted from the filters via centrifugation, quantified with a NanoDrop spectrophotometer, and then further digested by trypsin added at a ratio of 1:100. After incubation at 37 °C for 5 h, peptides were shock frozen in liquid nitrogen and lyophilized. Peptides (around 10 mg per Falcon tube) were re-suspended in 10 ml of ACN 80%, TFA 6%, and insoluble peptides were spun down by centrifugation at 100g for 1 min. Supernatants were moved into new 15-ml Falcon tubes, and samples were incubated twice with 50 mg of TiO$_2$ beads on a rotating wheel for 45 min. TiO$_2$ beads from all the enrichments were then pooled together and washed three times with 12 ml of ACN 80%, TFA 6% and three times with 12 ml of ACN 80%, TFA 0.1%. Beads were then re-suspended in 2 ml of ACN 80%, TFA 0.1%, transferred into 12 Empore-C8 StageTips (18), and washed once with ACN 80%, TFA 0.1%. Peptides were eluted from each StageTip with 200 μl of 60% NH$_4$OH (25% NH$_3$ solution in H$_2$O) in 40% ACN. The volume was reduced via SpeedVac to 10 μl to eliminate ACN and brought back up to 200 μl with 0.1% formic acid. Phosphorylation enriched peptides were pooled and purified with Sep-Pak tC$_{18}$ cartridges according to the manufacturer's instructions. The peptide concentration was determined using a NanoDrop spectrophotometer. The final concentration was brought to 400 ng/μl with 0.1% formic acid, and 5.5-μl aliquots were frozen at −20 °C.

LC-MS/MS Analysis—Online chromatography was performed with the Thermo Easy nLC ultra-high-pressure HPLC system (Thermo Fisher Scientific) coupled online to either an original Q Exactive or a Q Exactive HF with a NanoFlex source (Thermo Fisher Scientific). Ana-

lytical columns (50 cm long, 75-μm inner diameter) were packed in-house with ReproSil-Pur C$_{18}$ AQ 1.9-μm reversed phase resin (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) in buffer A (0.5% acetic acid). During on-line analysis the analytical column was placed in a column heater (Sonation GmbH, Biberach, Germany) regulated to a temperature of 55 °C. A peptide mixture of 2 μg dry weight was loaded onto the analytical column with buffer A at a maximum back-pressure of 980 bar (generally resulting in a flow rate of 450 nL/min) and separated with a linear gradient of 5% to 30% buffer B (80% ACN and 0.5% acetic acid) at a flow rate of 250 nL/min controlled by IntelliFlow technology over 90 min (generally at a back-pressure of around 500 bar). Due to the loading, lead-in, and washing steps, the total time for an LC-MS/MS run was about 40 to 50 min longer. Online quality control, including automated detection of large droplet formation, HPLC parameters, and acquisition-related computer status, was performed with SprayQc (8).

MS data were acquired using a data-dependent top-10 method for the Q Exactive and a top-15 method for the Q Exactive HF, dynamically choosing the most abundant not-yet-sequenced precursor ions from the survey scans (300–1650 Th). Sequencing was performed via higher energy collisional dissociation fragmentation with a target value of 1e5 ions determined with predictive automatic gain control (supplemental Fig. S6). Isolation of precursors was performed with a window of 3 Th for the Q Exactive and 1.4 Th for the Q Exactive HF, because of the latter's superior quadrupole (Fig. 3C). Survey scans were acquired at a resolution of 70,000 at *m/z* 200 on the Q Exactive and 60,000 at *m/z* 200 on the Q Exactive HF ("Results and Discussion"). Resolution for HCD spectra was set to 17,500 at *m/z* 200 with a maximum ion injection time of 120 ms on the Q Exactive and 15,000 at *m/z* 200 with maximum ion injection time of 25 ms on the Q Exactive HF (supplemental Fig. S3). The normalized collision energy was 25 for the Q Exactive and 27 for the Q Exactive HF (supplemental Fig. S2; this difference was due to different scaling functions in the instrument software). The "underfill ratio," specifying the minimum percentage of the target ion value likely to be reached at the maximum fill time, was defined as 10% (supplemental Fig. S5). Furthermore, the S-lens RF level was set at 60, which gave optimal transmission of the *m/z* region occupied by the peptides from our digest (supplemental Fig. S4). We excluded precursor ions with single, unassigned, or six and higher charge states from fragmentation selection (supplemental Fig. S2).

In the comparison between the Q Exactive and the Q Exactive HF, we noticed that there was a 1.5-times overestimation of the number of ions offered for fragmentation on the Q Exactive HF relative to the Q Exactive runs. To determine whether this had an additional effect on the performance of the instrument, we investigated performance on different target values. From this we found that the original Q Exactive target value of 1e5 ions remained optimal for the Q Exactive Plus and Q Exactive HF (supplemental Fig. S6).

Data Analysis—All data were analyzed with the MaxQuant proteomics data analysis workflow, version 1.4.0.6 (19), with the Andromeda search engine (20). The false discovery rate was set at 1% for protein, peptide spectrum match, and site decoy fraction levels. Peptides were required to have a minimum length of seven amino acids and a maximum mass of 4600 Da. MaxQuant was used to score fragmentation scans for identification based on a search with an allowed mass deviation of the precursor ion of up to 4.5 ppm after time-dependent mass calibration. The allowed fragment mass deviation was 20 ppm. Fragmentation spectra were searched by Andromeda in the International Protein Index human database (version 3.68; 87,061 entries) combined with 262 common contaminants (20). Enzyme specificity was set as C-terminal to arginine and lysine, also allowing cleavage at proline bonds and a maximum of two missed cleavages. We set carbamidomethylation of cysteine as a fixed mod-
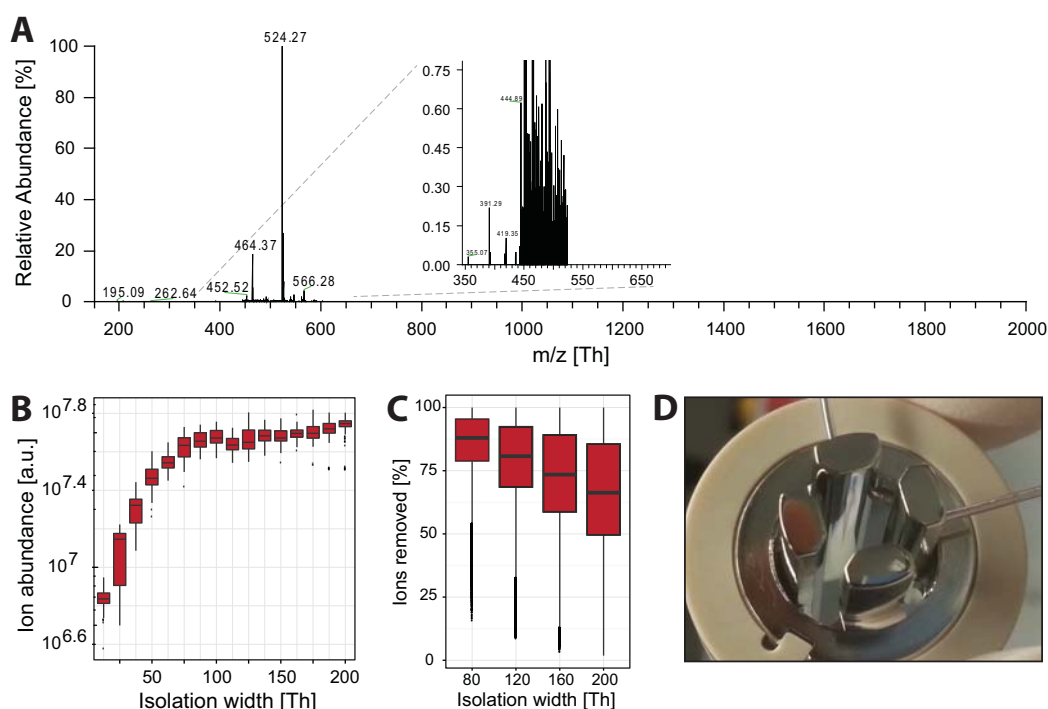
FIG. 2. **Performance details of the injection flatapole pre-mass filter.** *A*, the calmix spectrum when isolating MRFA (524.27 *m/z*) in an isolation window of 80 Th. The inset zooms in on the isolation window and its direct surroundings. *B*, isolation efficiency of the pre-mass filter over a range of isolation windows. *C*, effectiveness of the pre-mass filter over a range of isolation windows on a HeLa whole cell lysate. *D*, picture of the injection flatapole after 3 months of continuous measurements.

ification and N-terminal protein acetylation and oxidation (M) as variable modifications for the HeLa total cell lysates; additionally, Phospho (STY) was set as a variable modification for the phosphorylation enriched samples. Further downstream analysis of the results was performed with in-house-developed tools for the extraction of metadata from the mass spectrometry files based on MSFileReader (Thermo Fisher Scientific) and with the R scripting and statistical environment (21) using ggplot (22) for data visualization. The datasets used for analysis have been deposited at the ProteomeXchange Consortium via the PRIDE partner repository.

RESULTS AND DISCUSSION

The ion path during quadrupole isolation of the Q Exactive HF (as well as the Plus) has been updated with the objective of improving robustness and optimizing the ion transmission during quadrupole isolation in order to be able to meet the demand created by the increased speed of the ultra-high-field Orbitrap mass analyzer. Various hardware components making up the path are detailed in "Experimental Procedures." We start our discussion by investigating their behavior using the standard ESI Positive Ion Calibration Solution (Thermo Fisher Scientific) electrosprayed by direct infusion. We then characterize the performance of the instrument based on results from HeLa whole cell lysates measured via shotgun top-*N* methods.

Injection Flatapole as Pre-filter—We and others had found that due to the intense peptide ion beam, the Q Exactive could

be prone to contamination issues under heavy load that degraded its performance after a prolonged period of measurements. The leading cause was determined to be peptides excluded during mass selection coating the rods of the quadrupole. Therefore, the first update was made to the injection flatapole with the objective of making the instrument more robust. This component now acts as a low-resolution quadrupole capable of rough pre-filtering of the incoming ion beam before proper precursor selection in the analytical quadrupole, thereby providing an protective filter for contaminants. We tested the pre-filter on the tetra-peptide MRFA present in the electrospray ionization calmix solution (Fig. 2*A*). When the pre-filter was activated for the MRFA peak at 524.27 Th in an isolation window of 80 Th, a clean exclusion of peaks was achieved outside the selected isolation range. This is evident from the inset zoom-in to below 1% on the relative ion abundance scale. To determine the isolation width at which the pre-filter has optimal ion transmission, we started with a 10-Th selection window and increased it in 10-Th steps to 200 Th. The signal increased with increasing window size until 80 Th, after which hardly any increase in transmitted ions occurred (Fig. 2*B*). At this isolation width more than 90% of the used *m/z* scan-range (300–1650 Th) was excluded. By comparing the summed ion abundance of the signals from a HeLa total cell digest that were discarded when employing this
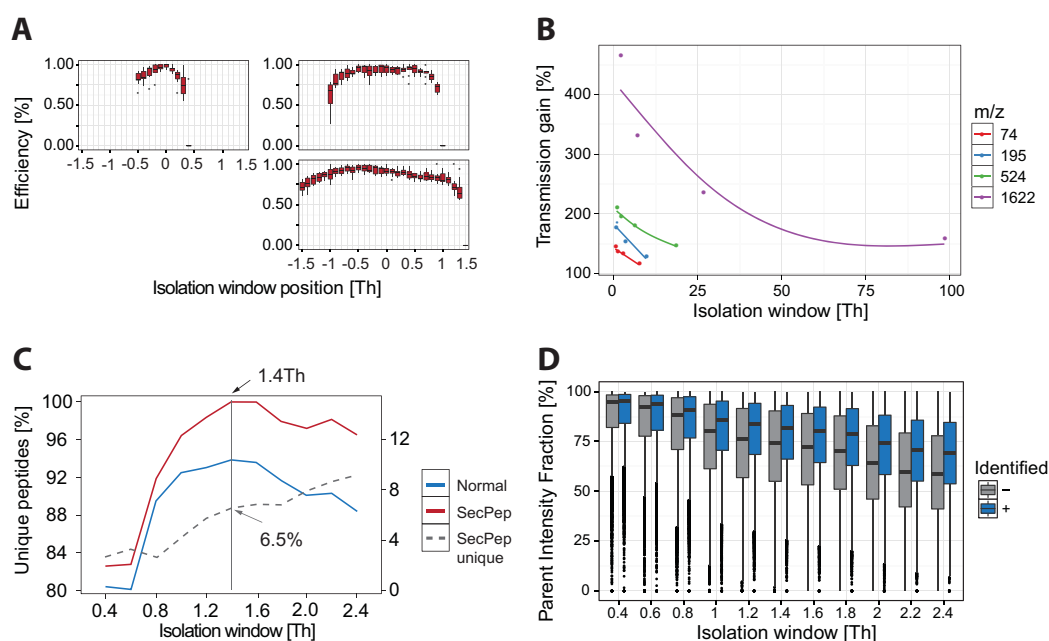
FIG. 3. **Transmission characteristics mass selection filter.** *A*, isolation window efficiency comparison between the Q Exactive and Q Exactive Plus using the complete set of calmix ions. *B*, isolation transmission efficiency comparison between the Q Exactive and the Q Exactive HF using the complete set of calmix ions. *C*, total unique peptides from a complex HeLa total cell lysate, sequenced with a range of isolation windows. The secondary *y*-axis denotes the contribution of the Andromeda option "second peptide" as a percentage (SecPep). *D*, isolation purity from a complex HeLa total cell lysate, sequenced with a range of isolation windows.

inject flatapole setting to the total ion current, we estimated that over 75% of the ion beam was excluded, for more than 75,000 of over 90,000 fragmentation scans (Fig. 2*C*). (Note that these numbers are somewhat dependent on the *m/z* range chosen.) With this low-resolution pre-filtering, a large portion of the selection load is moved to the robust inject flatapole. This, in turn, makes the task for the analytical quadrupole easier and should lead to less deposition on the selection quadrupole. The majority of non-precursor peptide ions are instead deposited on the injection flatapole (Fig. 2*D*), which is located in the outer cage and is readily accessible for cleaning.

Segmented Quadrupole—The original hyperbolic-rod quadrupole has been replaced by a segmented version capable of achieving more rectangular isolation efficiency over the complete isolation window (Fig. 3*A*). This is of relevance for acquisition strategies like SWATH (23) and co-isolation of SILAC partners (for example, in selected ion monitoring scans), which require rectangular isolation windows in order to generate accurate quantitative information at their edges. Relative to the quadrupole in the Q Exactive, we observed markedly improved isolation efficiency at the low-mass side of the isolation window. Additionally, the transmission of the quadrupole for narrow isolation windows has been improved. We measured this improved efficiency for the set of calmix ions and found an almost 2-fold increase (Fig. 2*B*). Note that

the figure also shows a transmission benefit for larger isolation windows, especially for high-*m/z* ions, that is due to the focusing effect of the exit segment of the quadrupole. This improvement roughly corresponds to the increase in ion current needed to support the doubling of the scan speed that the Orbitrap high-field analyzer is capable of (because the fill time corresponds to the transient time in fully parallel operation of the instrument).

Previously, the smallest isolation windows that we used were about 2.2 Th because transmission declined drastically below this value (24). As the segmented quadrupole also promised the ability to narrow isolation windows, we tested a range of window sizes on a complex HeLa whole cell lysate. We analyzed the data both with and without the "second peptide" option in Andromeda, which attempts to identify a second peptide from already identified MS2 scans after removal of the fragments associated with the first peptide (20). The Q Exactive HF performed optimally in terms of peptide identifications at an isolation window of 1.4 Th both with and without the second peptide option (Fig. 3*C*), whereas for the Q Exactive this value was 2.2 Th. The narrower isolation window also resulted in a reduced addition to the unique peptide sequences given by the second peptide option, because of the lower likelihood of co-isolated peptides that can be identified in this second step. At the optimal window of 1.4 Th this was 6.5%, down from 8% at the former optimal isolation
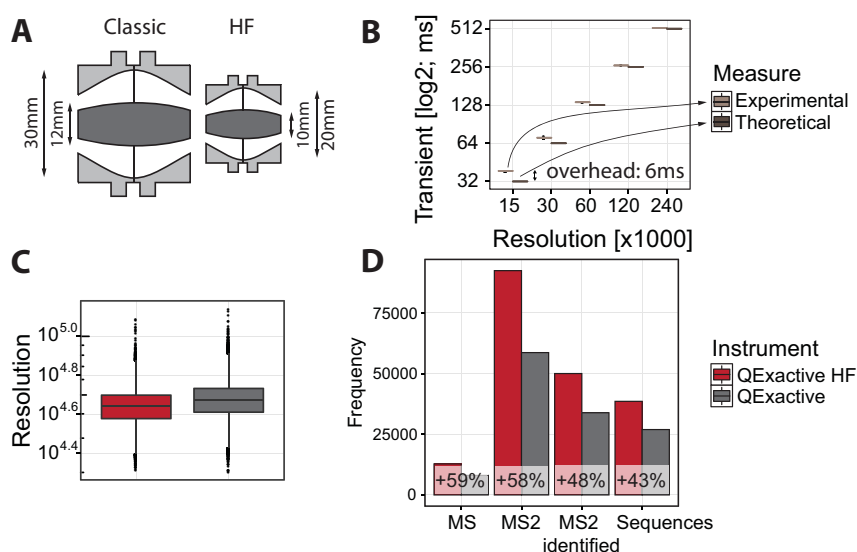
FIG. 4. **Performance of the Q Exactive HF with ultra-high-field Orbitrap analyzer.** *A*, comparison of dimensions of the standard (*left*) to the compact, ultra-high-field Orbitrap analyzer (*right*). *B*, comparison of the required Orbitrap transient times required for the different resolutions, together with an estimation of the overhead time required for each scan. *C*, measured resolution of peaks detected from a complex HeLa total cell lysate. *D*, scan statistics comparison between the Q Exactive and the Q Exactive Plus on HeLa total cell lysates.

window of 2.2 Th. We determined the purity of isolation by calculating the precursor ion fraction (24) for the same range of isolation windows. At the narrowest isolation window of 0.4 Th, the majority of isolations achieved greater than 80% purity (Fig. 3*D*). Such a low precursor ion fraction might to some degree be helpful for tandem mass tag labeling experiments (25, 26). For the optimal isolation window of 1.4 Th, the precursor ion fraction was slightly reduced to 70% in the large majority of cases. As this is still well within the range of tolerance for a clean and identifiable fragmentation spectrum, we adopted this isolation value as the default.

Performance of the Ultra-high-field Analyzer—The ultra-high-field Orbitrap analyzer consists of an outer barrel-like electrode (dark gray in Fig. 4*A*) of maximum radius R2 and a central spindle-like electrode (light gray) along the axis of maximum radius R1, with the outer electrode maintained at the virtual ground of the preamplifier and the central electrode at a voltage $-U_r$ ($U_r > 0$ for positive ions) (16). In a standard Orbitrap analyzer, R1 = 6 mm and R2 = 15 mm (5), whereas the ultra-high-field analyzer is more compact, with R1 = 5 mm and R2 = 10 mm (Fig. 4*A*) (*i.e.* the outer electrode is scaled down by a factor of 1.5). A decrease of the R2/R1 ratio from 2.5 to 2 allows an increase in the electric field and hence the detected frequency in addition to the scaling factor, thus bringing the total gain to about 1.8-fold. The smaller cell required an increase of the injection ion energy of about 1.4-fold for the same 5000-V voltage on the central electrode. Despite the increase in space charge density in the analyzer by a factor of $(1.5)^3 \approx 3.4$, the additional shielding provided by the relatively thicker central electrode keeps space-charge-

induced frequency shifts even slightly below those in the standard analyzer (27). A miniature electrostatic lens provides sharp spatial focusing of ions coming into the injection slot located on one of the injection electrodes. As ion packets from the C-trap enter the analyzer off-axis, axial oscillations are initiated without the need for any additional excitation and without a loss of synchronization with the moment of ejection from the C-trap. This facilitates the use of enhanced Fourier transform (28) for all modes of operation in the same way as in the preceding Q Exactive instrument.

The Q Exactive HF is preset to record spectral data at five distinct resolutions, namely, 15,000, 30,000, 60,000, 120,000, and 240,000 at *m/z* 200 Th. Each resolution corresponds to a transient time denoting the time spent on analyzing the ion population in the Orbitrap mass analyzer, respectively consisting of 32, 64, 128, 256, and 512 ms. These resolutions are slightly reduced for the ultra-high-field analyzer relative to the equivalent settings for the classic analyzer, as the total gain of the resolution at the same transient time is a factor of 1.8 rather than 2. The Q Exactive HF instrument additionally makes a higher maximum resolution of 240,000 available. By comparing the time difference between each consecutive scan for the ultra-high-field Orbitrap analyzer to the theoretical values for the transient times, we found that the overhead for each scan was between 6 and 14 ms (Fig. 4*B*). This constitutes an improvement over the 17-ms scan time overhead for the Q Exactive. When the increased transient speed is utilized to double the number of scans at roughly the same resolution, this reduction of the scan time overhead is required to keep the injection times at a reasonably high level when the instrument is running in fully parallel
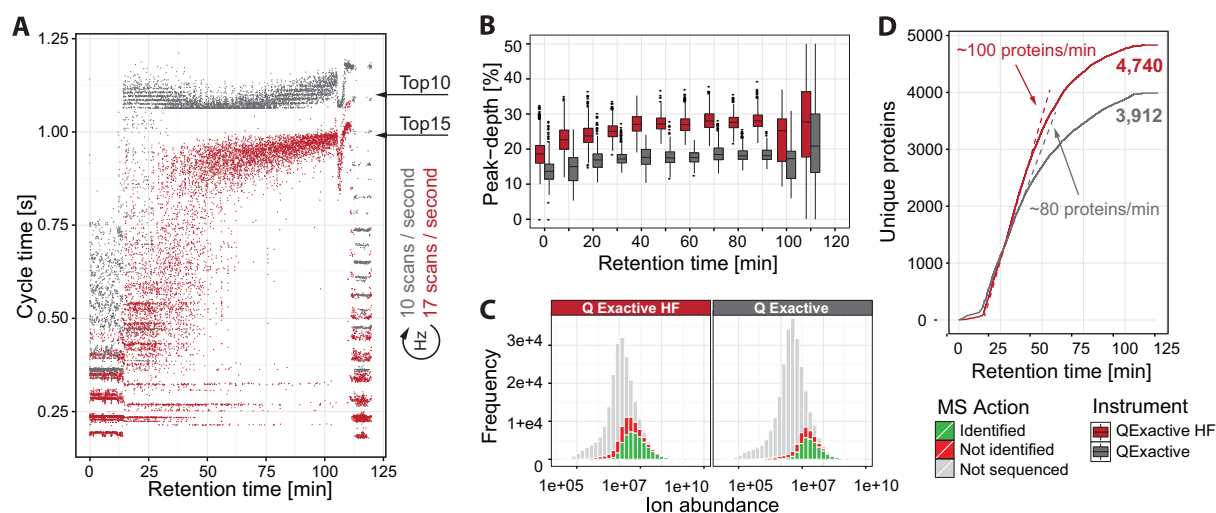
FIG. 5. **Performance of the Q Exactive HF with ultra-high-field Orbitrap analyzer.** *A*, cycle-time comparison for the Q Exactive and the Q Exactive HF on a complex HeLa total cell lysate. *B*, sequencing peak-depth comparison between the Q Exactive and the Q Exactive HF. *C*, comparison of the total number of visible peptides (cumulative green, red, and gray), sequenced peptides (cumulative green and red), and successfully identified peptides (green) between the Q Exactive and the Q Exactive HF. *D*, cumulatively identified proteins over the gradient comparison between the Q Exactive and the Q Exactive HF. The dotted lines denote a linear fit on the steepest part of the curve.

mode. For instance, for the 15,000 resolution setting the transient time is reduced from 64 to 32 ms; at a scan time overhead of 17 ms, more than half of the available time for the scan would have been lost without minimizing overhead. For shotgun analysis of complex mixtures, running in full parallel mode is desirable in terms of peptide identifications. We found the 15,000 resolution setting (corresponding to 32-ms transients) in combination with a maximum injection time of 25 ms to be best for complex mixtures such as total cell lysates (supplemental Fig. S3). To investigate the effect of the minor nominal reduction of instrument resolution, we measured the actual resolution of the identified peaks of a whole HeLa cell lysate. A slight drop in the resolution of those peaks was indeed observed (Fig. 4*C*); however, this is a minor effect, and we did not notice any impact on downstream data analysis. As described below, the increased scan speed indeed resulted in significantly increased numbers of peptide and protein identifications (Fig. 4*D*). We also note that one can double the MS resolution for the Q Exactive HF relative to that of the Q Exactive (120,000 *versus* 70,000) without detrimental consequences. This is because the transient time gained through acquisition at the lower resolution would be only 128 ms, which at 10% is an insignificant part of the total cycle.

Optimal Parameter Value Verification—After individually optimizing the parameters for the Q Exactive HF, we performed an independent verification of the optimal values for each of the parameters using a design-of-experiment approach (29, 30) using MODDE (Umetrics, Frankfurt, Germany). A benefit of such an approach is that it gives insight into parameters that have potentially interacting properties. For the analysis we chose the parameters that previously gave the greatest effect in our experiments, consisting of normalized collision energy

(20–40), S-lens rf level (40–80), and isolation window (0.4–3 Th). The range of the parameters was chosen to be large and center on the previously found optimal values. All RAW data were recorded based on the experimental design established by the software. From the results of this study we concluded that the previously described values are optimal and that in our set there were no significantly interacting parameters present from which a performance benefit could be gained. As the conclusion remained the same and a design-of-experiment approach can deal with sparse data, this can be an attractive study type for quickly optimizing LC-MS/MS metrics for different types of samples and chromatography separation (supplemental Fig. S10).

Q Exactive Performance for Total Cell Lysate Analysis—We next directly compared the Q Exactive HF to the Q Exactive using the standard top-10 method on a 2 h gradient for the Q Exactive, which has demonstrated effectiveness in addressing a wide variety of biological questions. We ensured a fair comparison by running the test within one day, using the same sample, HPLC, and analytical column on both systems. We wished to make use of the increased speed of the Q Exactive HF while still achieving a cycle time of around 1 s and occupying the instrument with close to the maximum number of fragmentation scans in each cycle. For this, we found a top-15 method to be the most effective. An example of the MS spectrum with the subsequent fragmentation spectra can be seen in supplemental Fig. S1. In this comparison, we achieved cycle times of 1.1 s on the Q Exactive and 0.9 s on the Q Exactive HF. The extra five scans per cycle at roughly the same cycle time translate into a speed increase from 10 scans per second to 17 scans per second (Fig. 5*A*). In a direct

comparison between the Q Exactive and the Q Exactive HF, we found that the higher speed delivered an increase of almost 60% more scans (variation over five runs $\pm$ 620 scans). At an identification success rate of 62% in both cases, this translated into 48% more MS2 identifications (variation over five runs $\pm$ 313 identifications) and 43% more unique sequences (variation over five runs $\pm$ 238 sequences) (Fig. 4*D*).

One prominent feature of the cycle-time plot is that the Q Exactive reached the full top *N* early on in the gradient, whereas the Q Exactive HF achieved the full top *N* much later in the gradient. We interpret this to mean that even in highly complex samples such as HeLa full cell lysates, the initial part of the gradient does not contain sufficient precursor ions that fulfill the criteria for selection for fragmentation at these very high sequencing speeds.

To investigate and compare the impact of the extra sequencing speed, we determined the achieved peak depth for each of the instruments. Here we define this value as the position of the precursor in the list of all visible precursors in the cycle where it is sequenced, sorted on descending ion abundance. For complex HeLa digests, we found that this list contained around 400 precursors (distilled from around 4000 individual peaks in the spectrum) at any given time during the linear part of the gradient (supplemental Fig. S7). The Q Exactive probed this list to a medium depth of 18%, or around 72 precursors per cycle, whereas the Q Exactive HF, with its enhanced speed, achieved a depth of 27%, or around 108 isotopes per cycle (Fig. 5*B*). This additional peak depth was indeed translated into more fragmented precursors, which were increased by 53% from 48,449 to 74,383. As both instruments exhibited a sequencing success rate of more than 50% over the entire gradient, this translated into an increase of 43%, from 27,256 to 39,119 unique peptide sequences (Fig. 5*C*). There was a slight reduction in the number of visible peptides (cumulative green, red, and gray population), which we attribute to the lower resolution of the full scans for the Q Exactive HF. From the histogram, however, it is evident that the instrument sequenced down to lower abundance peptides (cumulative green and red population) and is capable of successfully identifying a greater proportion of available peptides (green population). These additionally sequenced peptides contribute to new protein identifications, as opposed to only extending the sequence coverage of already accessed proteins. This is evident from a 21% increase in identified proteins, from 3912 to 4740. Thus, the sequencing speed enabled by the ultra-high-field Orbitrap analyzer allows identification of a greater percentage of total detectable precursors, but a majority of unfragmented peptides remains. This is not entirely due to their low abundance, as many peptides in the gray population have a signal similar to those of fragmented and successfully identified peptides. This suggests that further increases in sequencing speed would still be useful for particularly congested parts of the gradient. Plotting

the cumulative number of proteins over the gradient revealed that the Q Exactive HF improved the maximum protein sequencing rate by 25%, from 80 to 100 proteins per minute (Fig. 5*D*).

Further investigations into the dynamic range of the identified peptides revealed that both instruments reliably sequenced over 3 orders of magnitude, indicating that the performance increase of the Q Exactive HF can be attributed to its ability to sequence more peptides in the busy regions of chromatography. This additional speed is also responsible for a higher inter-replicate reproducibility at the peptide identification level, where the instrument achieves better reproducibility for lower peptide abundances (supplemental Fig. S8). To investigate the Q Exactive HF performance on different gradient lengths and to provide an indication of optimal performance for the instrument, we additionally ran a gradient titration series on the standard HeLa samples, which revealed that gradients over 4 h did not further improve identification performance for our sample. Based on the increase per time unit, we concluded that a 150-min gradient represents the optimal length. The trend in our and other laboratories to measure over 4-h or longer gradients can now to some extent be reversed (supplemental Fig. S9).

As we observed that the Q Exactive HF did not routinely achieve the full top *N*, we hypothesized that the peak depth was limited by chromatography rather than the sequencing speed of the mass spectrometer. To address this question, we employed DMSO as a dopant in the mobile phase buffers, as it had been reported to significantly increase the ion flux and consequently the performance of shotgun proteomics experiments (31). On the standard 2 h gradient, peptide identifications increased by 13%, from 39,119 to 44,446, and protein identification increased by 16%, from 4740 to 5492 (Fig. 6*A*). Even with this addition of potentially sequenceable peptides, the instrument still did not reach the full top *N* per cycle, meaning that further increases in ion abundance would be beneficial. The achieved protein sequencing rate over the actual gradient (subtracting the lead-in time of 5 min for the buffers to arrive and a washout phase of 15 min) was 55 proteins per minute for these optimized conditions. When comparing the copy numbers for HeLa cells to the detected proteins (32), we found that the instrument was capable of successfully sequencing in an estimated protein dynamic range from 90 million down to ∼115 copies per cell (Fig. 6*B*), enabling the analysis of highly complex samples down into the transcription factor range in reasonable time. Although the dimethyl sulfoxide dopant in our experiments had clear benefits in terms of performance, we do not routinely use it because it represents a potential contamination source for the quadrupole rods, which cannot be resolved through the incorporation of the selecting inject flatapole. Further studies are clearly needed to confirm its effects on instrument robustness.

Application to Phosphoproteomics—Given the observed increase in performance for the whole cell lysates, we reasoned that the Q Exactive HF might be capable of improved per-

FIG. 6. **Optimized chromatography performance.** *A*, cumulative proteins over the gradient. *B*, copy numbers per cell for the detected proteins.

formance on sample types requiring high mass spectrometric sensitivity as well. In order to test this, we measured an unstimulated HeLa lysate enriched for phosphorylated peptides with both instruments. Phosphopeptides are generally of low abundance, and therefore it is advantageous to allow for longer ion injection times for the fragmentation scans to accumulate a sufficient amount of ions for a successful sequencing event (33). Here we chose a maximum of 111 ms for the Q Exactive and 45 ms for the Q Exactive HF, which took the 17-ms scan time overhead of the original Q Exactive into account (leaving the Q Exactive HF at a slight disadvantage). Fig. 7*A* illustrates the analysis on both instruments for the same phosphorylated peptide. Even though in both cases the maximum allowed ion injection time was reached—meaning that not enough ions were present to fulfill the request of 1e5 ions for the higher energy collisional dissociation scan— almost full sequence coverage was achieved in both cases. The Q Exactive HF, however, achieved this coverage in less than half the ion injection time. We investigated whether the automatic gain control target could still be achieved at these restricted maximum injection times for all the MS2 scans and found that it could in the majority of cases for both instruments. As in the above example, most of the MS2 events for the classic instrument required almost double the time in this experiment (supplemental Fig. S11). The longer ion injection time in the sensitive method already allows the Q Exactive to probe deeper into the list of visible precursors, from a median of 18% for the normal method to 23% for the sensitive method. However, the additional speed of the Q Exactive HF allowed it to go even further, from a median of 27% for the normal method to 33% for the sensitive method (Fig. 7*B*). Overall, the increased sequencing speed in combination with
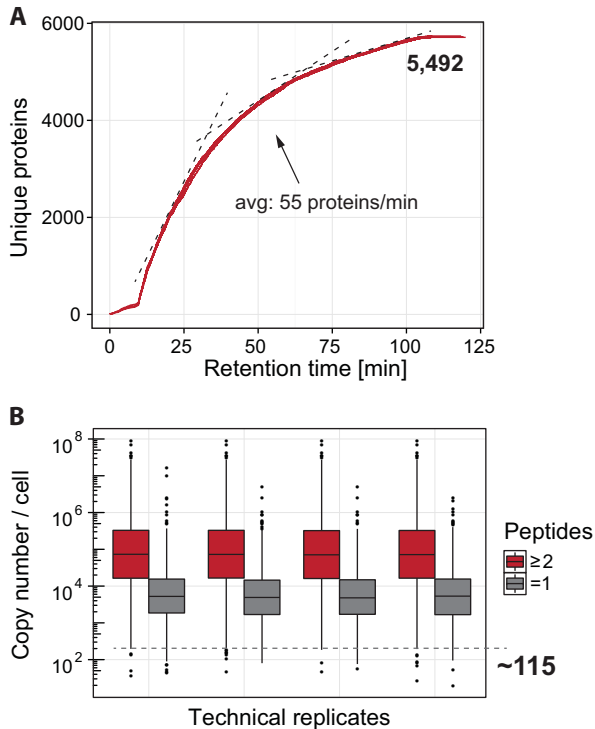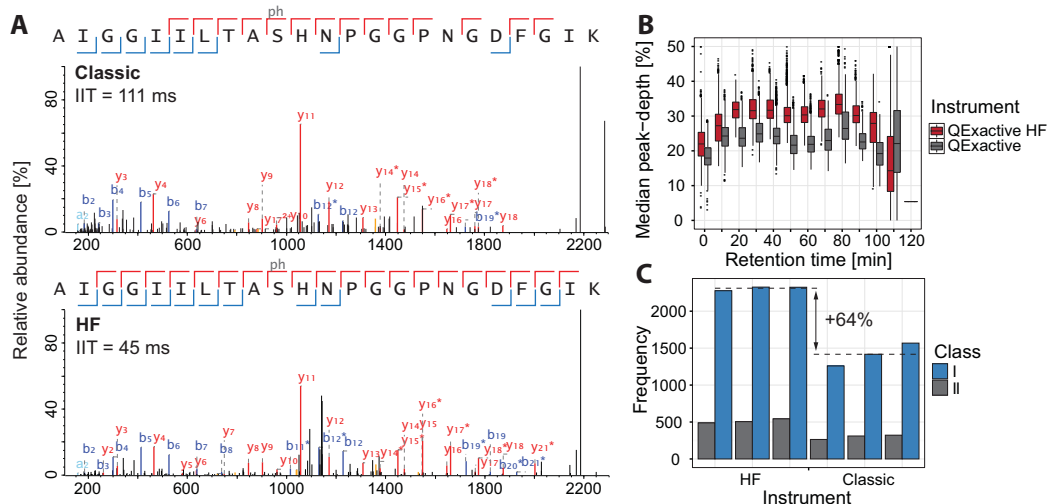

FIG. 7. **Phosphorylation enriched samples.** *A*, example fragmentation spectra of the same phosphorylated peptide identified on the Q Exactive (*top*) and on the Q Exactive HF (*bottom*). In both cases the maximum allowed ion injection time was reached. *B*, comparison of the achieved peak-depth on the Q Exactive and the Q Exactive HF. *C*, comparison of the successfully sequenced Class I phosphorylation sites.

the improved quadrupole transmission translated into 64% more class I phosphorylation sites identified in a single run in this particular experimental situation (Fig. 7C). However, in experimental situations where the phosphopeptide amounts necessitated longer fill times, we have observed much smaller gains.

Conclusions and Outlook—Here we have described the next iteration of the Exactive family of mass spectrometers, the Q Exactive Plus and the Q Exactive HF, and investigated its performance on analytical standards and on complex peptide mixtures for shotgun proteomics using our standard 90-min gradient. The ultra-high-field Orbitrap analyzer doubles the sequencing speed at the same resolution, for which the improved ion transmission characteristics of the segmented quadrupole at least partially provide the necessary increase in precursor ion abundance. The higher speed translates well to actual shotgun proteomics improvements, as we observed increases of more than 40% in unique peptide sequences and more than 20% in proteins relative to the previous generation on our standard gradient with the HeLa cell lysate.

We observed that the mass spectrometer did not fully make use of its potential sequencing speed, as there were insufficient ions in parts of the gradient to satisfy the requirements for attempting a sequencing event. In order for this to be resolved, optimizations such as dopant-enhanced mobile phases (31), higher sensitivity analytical columns (34), or brighter ion sources (35) will be required to increase the precursor ion abundance. In applications with very complex samples containing a sufficient number of sequenceable precursors, we anticipate that researchers will be able to boost the instrument productivity almost 2-fold while still achieving the same sequencing depth as before. With such a large reduction in analysis time, any overhead between LC-MS/MS runs becomes increasingly undesirable. For example, currently the loading times on our single analytical column setup (50 cm; C_{18}) are around 30 min. During this time the mass spectrometer is not recording useful data, which is still acceptable for an analysis time of 2 h given that this loss corresponds to two runs per day. At a 1-h analysis time, however, an overhead of 30 min results in a loss of almost eight runs per day. Thus optimized chromatography setups, such as pre-columns that do not degrade resolution or double analytical columns, will be needed in order for the instrument to be used most effectively.

In conclusion, this new generation of quadrupole Orbitrap instruments is designed for significantly greater robustness, and our experiments confirmed that the majority of undesired ions were confined to the front part of the instrument. We likewise observed significantly enhanced selection characteristics of the new segmented quadrupole, which should be useful in many proteomic experiments. Finally, and most important, the ultra-high-field Orbitrap analyzer in the Q Exactive HF routinely provided doubled resolution in MS scans without downsides and doubled the potential sequencing speed in MS/MS mode.

REFERENCES

1. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422**, 198–207
2. Altelaar, A. F., and Heck, A. J. (2012) Trends in ultrasensitive proteomics. *Curr. Opin. Chem. Biol.* **16**, 206–213
3. Nagaraj, N., Wisniewski, J. R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Paabo, S., and Mann, M. (2011) Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* **7**, 548
4. Geiger, T., Wehner, A., Schaab, C., Cox, J., and Mann, M. (2012) Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol. Cell. Proteomics* M111.014050
5. Beck, M., Schmidt, A., Malmstroem, J., Claassen, M., Ori, A., Szymborska, A., Herzog, F., Rinner, O., Ellenberg, J., and Aebersold, R. (2011) The quantitative proteome of a human cell line. *Mol. Syst. Biol.* **7,** 549
6. Nesvizhskii, A. I., and Aebersold, R. (2005) Interpretation of shotgun proteomic data: the protein inference problem. *Mol. Cell. Proteomics* **4,** 1419–1440
7. Pichler, P., Mazanek, M., Dusberger, F., Weilnböck, L., Huber, C. G., Stingl, C., Luider, T. M., Straube, W. L., Kocher, T., and Mechtler, K. (2012) SIMPATIQCO: a server-based software suite which facilitates monitoring the time course of LC-MS performance metrics on Orbitrap instruments. *J. Proteome Res.* **11,** 5540–5547
8. Scheltema, R. A., and Mann, M. (2012) SprayQc: a real-time LC-MS/MS quality monitoring system to maximize uptime using off the shelf components. *J. Proteome Res.* **11,** 3458–3466
9. Nagaraj, N., Kulak, N. A., Cox, J., Neuhaus, N., Mayr, K., Hoerning, O., Vorm, O., and Mann, M. (2011) Systems-wide perturbation analysis with near complete coverage of the yeast proteome by single-shot UHPLC runs on a bench-top Orbitrap. *Mol. Cell. Proteomics,* M111.013722
10. Kocher, T., Swart, R., and Mechtler, K. (2011) Ultra-high-pressure RPLC hyphenated to an LTQ-Orbitrap Velos reveals a linear relation between peak capacity and number of identified peptides. *Anal. Chem.* **83,** 2699–2704
11. Thakur, S. S., Geiger, T., Chatterjee, B., Bandilla, P., Frohlich, F., Cox, J., and Mann, M. (2011) Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. *Mol. Cell. Proteomics* **10,** M110.003699
12. Zubarev, R. A., and Makarov, A. A. (2013) Orbitrap mass spectrometry. *Anal. Chem.* **85,** 5288–5296
13. Olsen, J. V., Schwartz, J. C., Griep-Raming, J., Nielsen, M. L., Damoc, E., Denisov, E., Lange, O., Remes, P., Taylor, D., Splendore, M., Wouters, E. R., Senko, M., Makarov, A., Mann, M., and Horning, S. (2009) A dual pressure linear ion trap orbitrap instrument with very high sequencing speed. *Mol. Cell. Proteomics* **8,** 2759–2769
14. Michalski, A., Damoc, E., Hauschild, J. P., Lange, O., Wieghaus, A., Ma-

karov, A., Nagaraj, N., Cox, J., Mann, M., and Horning, S. (2011) Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol. Cell. Proteomics* **10,** M111.011015

15. Gallien, S., Duriez, E., Crone, C., Kellmann, M., Moehring, T., and Domon, B. (2012) Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer. *Mol. Cell. Proteomics* **11,** 1709–1723

16. Michalski, A., Damoc, E., Lange, O., Denisov, E., Nolting, D., Mueller, M., Viner, R., Schwartz, J., Remes, P., Belford, M., Dunyach, J. J., Cox, J., Horning, S., Mann, M., and Makarov, A. (2011) 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

17. Wisniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nat. Methods* **6,** 359–362

18. Rappsilber, J., Ishihama, Y., and Mann, M. (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75,** 663–670

19. 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

20. Cox, J., Neuhauser, N., Michalski, A., Scheltema, 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

21. Ihaka, R., and Gentleman, R. (1996) R: a language for data analysis and graphics. *J. Comput. Graph. Stat.* 299–314

22. Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis*, Springer, New York volume: 5

23. Gillet, L. C., Navarro, P., Tate, S., Roest, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012) Targeted data extraction of the MS/MS spectra generated by data independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11,** O111.016717

24. Michalski, A., Cox, J., and Mann, M. (2011) More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *J. Proteome Res.* **10,** 1785–1793

25. Savitski, M. M., Sweetman, G., Askenazi, M., Marto, J. A., Lang, M., Zinn, N., and Bantscheff, M. (2011) Delayed fragmentation and optimized isolation width settings for improvement of protein identification and accuracy of isobaric mass tag quantification on Orbitrap-type mass spectrometers. *Anal. Chem.* **83,** 8959–8967

26. Thompson, A., Schafer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A. K., and Hamon, C. (2003) Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **75,** 1895–1904

27. Makarov, A., Denisov, E., and Lange, O. (2009) Performance evaluation of a high-field Orbitrap mass analyzer. *J. Am. Soc. Mass Spectrom.* **20,** 1391–1396

28. Lange, O., Damoc, E., Wieghaus, A., and Makarov, A. (2014) Enhanced Fourier transform for Orbitrap mass spectrometry. *Int. J. Mass Spectrom.* **369,** 16–22

29. Montgomery, D. C. (2000) *Design and Analysis of Experiments*, 5th Ed., Wiley Hoboken, New Jersey

30. Randall, S. M., Cardasis, H. L., and Muddiman, D. C. (2013) Factorial experimental designs elucidate significant variables affecting data acquisition on a quadrupole Orbitrap mass spectrometer. *J. Am. Soc. Mass Spectrom.* **24,** 1501–1512

31. Hahne, H., Pachl, F., Ruprecht, B., Maier, S. K., Klaeger, S., Helm, D., Medard, G., Wilm, M., Lemeer, S., and Kuster, B. (2013) DMSO enhances electrospray response, boosting sensitivity of proteomic experiments. *Nat. Methods* **10,** 989–991

32. Wiœniewski, J. R., Hein, M. Y., Cox, J., and Mann, M. (2014) A 'proteomic ruler' for protein copy number and concentration estimation without spike-in standards. *Mol. Cell. Proteomics* M113.037309

33. Kelstrup, C. D., Young, C., Lavallee, R., Nielsen, M. L., and Olsen, J. V. (2012) Optimized fast and sensitive acquisition methods for shotgun proteomics on a quadrupole Orbitrap mass spectrometer. *J. Proteome Res*. **11,** 3487–3497

34. Zhou, F., Lu, Y., Ficarro, S. B., Webber, J. T., and Marto, J. A. (2012) Nanoflow low pressure high peak capacity single dimension LC-MS/MS platform for high-throughput, in-depth analysis of mammalian proteomes. *Anal. Chem.* **84,** 5133–5139

35. Belov, M. E., Damoc, E., Denisov, E., Compton, P. D., Horning, S., Makarov, A. A., and Kelleher, N. L. (2013) From protein complexes to subunit backbone fragments: a multi-stage approach to native mass spectrometry. *Anal. Chem.* **85,** 11163–11173

Concluding Remarks and Outlook

Proteomics Tools to Investigate Neurodegeneration

Neurodegenerative diseases (NDs) are an increasing burden for our aging society. A wide spectrum of these pathologies is characterized by progressive neuronal death causing functional impairment of the affected brain regions. Despite decades of intense research, neither effective treatments nor preventions are available, rendering neurodegeneration one of the major causes of morbidity and mortality. In this thesis, proteomics tools were developed and employed to provide novel insights into the pathology of neurodegeneration.

Mass spectrometry based proteomics provides quantitative measures for protein abundance and allows us to monitor the physiology of tissues and cells as well as changes during the development of pathologies. *Article 1* demonstrates a paradigm example of how the physiological properties of primary neurons and neuronal model cell lines can be assessed with proteomics. We evaluated different *in vitro* cell systems for their cellular properties in comparison to primary motoneurons. The motoneuronal model cell lines investigated are widely used to study neurodegeneration. Surprisingly, our proteomics results indicated that these cell lines significantly differ from primary motoneurons both on the global pathway and individual protein levels. The neuronal lineage of the motoneuron model cell lines only became evident in comparison to non-neuronal cell lines. With the results from *article 1* it is possible to infer which pathways and functionalities are preserved in the cellular model systems, allowing data driven decision which cellular model system to use for specific questions. In addition, our approach illustrates how proteomics can be used as a tool to probe almost the entire proteome for physiological and pathological changes in a few hours of measuring time.

Toxic aggregates are a hallmark of NDs. However, they differ in many aspects such as size, localization, associated pathology and composition. While the first three properties can be investigated with standard biochemical approaches, proteomics is the only technology that can assess the composition of aggregates in an unbiased and quantitative fashion. In *article 2* we analyzed the proteomic composition of SDS insoluble poly-dipeptide repeat aggregates (pDRP) that are associated with a c9orf72 hexanucleotide repeat expansion, the most common genetic cause for ALS and FTD [48]. We observed a strong association of the UPS with the pDPR aggregates and discovered UNC119, which is quantitatively sequestered into the aggregates, as a novel player in ALS and FTD pathologies. In primary neurons, overexpression of UNC119 is sufficient to rescue the pDRP phenotype partially, whereas its deficiency compromises neuronal survival. In post mortem brain

tissue, sequestration of UNC119 correlates better with the degree of degeneration than the pDPR aggregate formation itself. Further investigation on these effects as well as on the contribution of other proteins that we determined to be highly enriched in the aggregates *in vivo* are ongoing.

Important questions in neurodegenerative research include how the toxicity of aggregates can be modified and which natural mechanisms have evolved to cope with proteostatic stress. We employed yeast as a model system in *article 3* to address these questions. In yeast, the toxicity of N-terminal fragments of huntingtin containing a polyQ stretch (103Q-GFP) can be modulated in genetic screens, providing a versatile screen for toxicity modifying mechanisms. We assessed how proteins that suppress 103Q-GFP toxicity, like the non-essential transcription factor GTS1, affect aggregate composition. We analyzed the HTT interactome and discovered that the composition of the aggregates significantly changes upon pertubation. In the presence of overexpressed GTS1 the association of HTT with prion-domain containing proteins is decreased, whereas chaperone association is increased. Moreover, we showed that overexpression of the prion-like polyQA domain of GTS1 alone is sufficient to suppress 103Q-GFP toxicity and to alter aggregate composition. Yeast expresses a set of proteins with a prion-like behavior. For yeast, those prion proteins do not necessarily exert toxicity and our findings support the proposed beneficial effects under certain circumstances [120]. Furthermore, proteins with prion like domains are not unique to yeast; and indeed, many proteins of higher eukaryotes contain prion like domains [121] and might partially contribute to proteostasis with related mechanisms.

Protein aggregates in NDs differ in general features such as localization, shape and co-aggregating proteins. This suggests that the toxicity of protein aggregates is not generally mediated via a simple mechanism like sequestration of a small number of proteins but rather involves a multicomponent network. Such a network would include various deleterious effects such as protein aggregates and impaired proteostasis, disturbed RNA metabolism, non–cell autonomous toxicity and neuroinflammation. In line with the 'multiple hit hypothesis' as applied to neurodegeneration [44,229], the different neurodegenerative pathologies might be linked in this complex network, of which we only have a partial understanding so far. Unbiased quantitative approaches such as proteomics could help to gain comprehensive insights into this network, identifying the missing links and shedding light on the pathological hubs. From a proteomics point of view, we can address three major aspects: (1) The physiology of neurons and glial cells and how the complete proteomes alter during the pathology (2) A comparison of sub proteomes of neurons and glia cells in the context of various NDs and (3) Characterization of different neurodegeneration associated aggregates.

(1) Characterizing Complete Proteomes: Most non-neuronal cells, but even some neurons such as the oculomotor nerve in ALS, are less prone to degeneration. An unbiased characterization of primary cells or post mortem tissue of different neurons by mass spectrometry will help to

understand why certain neurons are particularly susceptible. Besides primary cells or post mortem tissue dissections, patient derived induced pluripotent stem cells (iPSC) differentiated into individual neurons could provide powerful disease models for proteomics analysis. Moreover, a longitudinal proteome analysis during a differentiation process of iPSC towards neurons and the onset of a phenotype would give detailed insights into pathological cascades and causality. As a control, one could use cells derived from healthy family members. To rescue from specific mutations one could perform genome editing with CRISPR-Cas9. A proof of concept for genome editing in the context of neurodegeneration was published in 2011 using zinc finger nucleases to generate isogenic human iPSC [230]. In addition to the analysis of neurons, non-cell autonomous toxicity in NDs can be investigated. Glia cells are important to maintain neuronal homeostasis and regulate inflammatory processes [18,30]. On the one hand, the specific neuronal and glial morphology and interconnection renders an analysis of individual cell types from this tissue almost impossible. On the other hand, iPSC derived and primary cells lack the cellular context that might play a pivotal role in neurodegeneration. Therefore, a combined analysis of both, tissue and primary / iPS cells is likely necessary for comprehensive understanding of the pathology. Expanding the proteomics characterization of primary cells, tissues and patient derived iPSC to multiple pathologies would allow discrimination of common and unique neurodegenerative features on a global scale. As different NDs affect different parts of the brain, healthy controls are required to normalize for the tissue bias.

(2) Characterizing Sub Proteomes: Detailed mechanistic insights in pathogenic cellular mechanisms can be gained by focusing on subcellular proteomes. Disturbance of neuronal transport processes could be investigated by focusing on the vesicle cargo [136] and analyzing the proteomic content of axons. The later can be achieved by using specifically designed growth chambers that allow separating the axons from the cell body [231]. Proteins secreted by neurons and glial cells are especially interesting to characterize non-cell autonomous toxicity. The inflammatory response of microglia to extracellular aggregates or the changed secretion pattern of neurons and glial cells harboring neurodegeneration associated mutations could be investigated using *in vitro* systems like primary cells or iPSC.

(3) Characterizing Protein Misfolding and Aggregation: *Article 2* and *3* illustrate approaches in two model systems that provided novel insight to ALS / FTD and prion pathology. An unbiased resource for the composition of many different aggregates does currently not exist, however it is known that inclusions show specificity for only a subset of aggregation prone proteins. For instance, some ALS associated aggregates are TDP positive while others are TDP negative [232].

Generally, this could have many reasons including different cellular localization, altered abundance of proteins, and might also point to the requirement of specific interfaces for co-aggregation [101].

This compatibility for transmission and amplification of 'misfolding information' shows intriguing conceptual parallels to classic prions, which propagate 'folding information' in an (auto-)catalytic fashion and are transmissible between organisms. A growing body of evidence points towards intercellular transmission and prion like behavior for proteins not previously thought to be involved in prion pathologies. Prominent examples of proteins that can propagate misfolding information to other proteins are tau, alpha-synuclein, SOD1, HTT and TDP-43 [34,36,233,234]. Although propagation of these proteins within one organism has been described, there is no evidence of natural transmission across species or even between individuals of one species [34]. However, this remaining difference may solely reflect the protein stability and ability to cross the cellular barriers into the central nervous system. Which sequence features and conditions render proteins compatible for coaggregation and how generic and specific epitopes contribute to that process could be assessed with comprehensive data on the 'aggregomes' of many neurodegenerative diseases. Such data would facilitate the understanding of the biophysical principles of misfolding, propagation of misfolding information and protein aggregation in diseases.

Article 3 provides an example of how the toxicity of aggregates correlates with the change in aggregate structure and composition. Quantifying the proteomic composition of aggregates in different pathologies would allow to investigate whether cells attempt to alter aggregate composition in general in response to proteostasis stress. If so, protein aggregates should contain the following groups of proteins. First, pathologically irrelevant binders consisting of proteins that stick to aggregates with little impact on the cell viability. Second, proteins that are sequestered causing loss of function toxicity and third, proteins that reduce the toxicity of the aggregates. Correlating the aggregate content with toxicity and the soluble protein content in established model systems such as primary neurons would help to discriminate these groups: if the protein abundance in the soluble fraction anti correlates with toxicity, this is consistent with a loss of function due to sequestration. However, if the abundance in the aggregates correlate with improved cell survival, it suggests that the protein acts as suppressor of aggregate toxicity. Unfortunately, this approach is complicated by the fact that toxicity can also be exerted by soluble oligomers of misfolded proteins. Therefore, a biochemical characterization and purification of the different species is imperative.

Technological Advances in Proteomics

Despite the huge technological improvements over the last decade, a need for improved MS performance will continue for many years. Swissprot currently lists 20,198 proteins for *H.sapiens* (March 2015). State of the art mass spectrometry can quantify more than 10,000 proteins [129] and cells cannot possibly express all possible proteins at the same time. Yet it is clear that current sample preparation and MS workflows are far from sensitive enough to achieve complete sequence coverage of all the expressed proteins. Over 90,000 isoforms are predicted in the human proteome [235]. Those are often particularly difficult to quantify as they can have very subtle sequence differences to each other. In conjunction with all PTMs, the potential complexity of the proteome is many orders of magnitude higher than the current capacity of MS technology. Some applications are particularly challenging due to very limited material or extraordinarily high dynamic ranges. Extremely high sensitivity, for example, would allow the analysis of the axonal proteome via microdisection. This could be used to investigate neuronal-oligodendrocyte crosstalk at the myelin interface, which is important for neuron homeostasis and is impaired in neurodegeneration [236]. Dynamic range is an issue, for example, when one wishes to investigate patient serum to diagnose diseases. In serum, less than 0.1 % of the proteins account for 95 % of proteome mass [237]. To meet these challenges, further developments of the MS workflow are urgently required.

Sample preparation enabling efficient and unbiased isolation of all proteins is the first step in proteome analysis. Many efforts have been made to decrease sample loss during proteolytic digestion and to increase the efficiency of protein isolation for low amount samples [238,239]. Although trypsin and LysC have proven to be invaluable proteases for most proteomics approaches, supplementing these by other proteolytic strategies can be useful. Combination of proteases can increase the sequence coverage by generating more peptides that can potentially be detected in the MS. In *article 4* we employed various proteases and partial digestion to characterize a so far unknown third proteolytic pathway in Alzheimer's disease and consequently identified a novel toxic peptide termed Aη-α. To increase the sequence coverage further one could employ less specific proteases like proteinase K. However, this approach would reduce the intensity of individual peptides roughly by the amount of additional peptides, compromising sensitivity. In addition, the expanded search space increases the computational load in down-stream data analysis. Apart from online liquid chromatography separation, the complexity of protein or peptide mixtures can be reduced by various fractionation methods usually based on ion exchange, hydrophobicity or isoelectric focusing [212,240-243]. Employing extensive fractionation of the sample can provide data with a higher dynamic range. However, more input material and measuring time is required and the additional processing steps can reduce quantitative accuracy. Ultimately, the best approach depends on the biological question that

is asked. For instance, high throughput screenings might rather benefit from fast and highly accurate data collection while a complete characterization of a cellular system may instead require extensive fractionation and a very deep proteome.

In the future, several MS developments will increase the quality of proteomics data and range of applications. More than ten years ago, the introduction of Orbitrap analyzers constituted a breakthrough for proteomics and since then it has contributed to the increasing popularity and applications of MS. The newest generation of hybrid Orbitrap MS evaluated in *article 6* almost doubled performance compared to its predecessor, featuring twice the resolution or sequencing speed with the same gradient length. The sensitivity was significantly improved by higher transmission efficiency in the segmented quadrupole. In addition, a pre-mass filter increased robustness of the machine significantly. Especially for clinical applications were speed and robustness are critical, the new generation quadrupole Orbitraps is a leap forward. The performance boost of the current generation was partially achieved by decreasing the size of the Orbitrap analyzer. However, for future iterations of hybrid Orbitrap MS reducing the cell size further will eventually decrease the ion capacity and increase space charge effects, decreasing accuracy and dynamic range. In addition, the finite ion capacity of the C-trap has always been a bottleneck for the achievable dynamic range. The beam type TOF mass analyzer, in contrast, has no ion capacity limitation, potentially allowing a very high dynamic range. Moreover, the TOF has the advantage to scale performance with drift path-length of ions in contrast to the reduction required for the Orbitrap. In *article 5* we benchmarked the next generation hybrid QTOF instrument and showed significant performance improvement for proteomics applications. With increasing resolution, practically unlimited mass range and higher dynamic range TOF technology has the capacity to potentially outcompete the Orbitrap technology in the future.

A potentially very useful addition to the current hybrid mass spectrometer would be ion mobility separation (IMS). IMS allows separating ions by shape and interaction with the gas phase and thereby adds orthogonal information to the molecule [244]. Especially complex mixtures could benefit from this additional separation and the shape-information may be useful to infer ion identity and to match peptides only identified in one samples to other samples. An early example is the investigation of the plasma proteome using IMS in combination with TOF [245]. The capacity of ion mobility to separate by shape may also become very useful to investigate glycosylation patterns and discriminate carbohydrate isomers. These are specifically interesting in determining immune-relevant antigens or in profiling carbohydrate dependent drug binding capacity of proteins and peptides.

To date, two different acquisition modes are employed in shotgun proteomics. In DDA ions are selected for fragmentation depending on the MS^1 information. This is a hypothesis free and very efficient acquisition as signals that are more likely to provide sufficient information for identification are preferred. The partially stochastic nature of DDA, however, makes it prone to undersampling, which adds variation to the data. The high sequencing speed of the current Q Exactive and QTOF generations make undersampling a minor issue (*article 5* and *6*), however, higher sensitivity or shorter gradients will increase the 'need for speed'. In contrast to DDA, SRM / MRM are a hypothesis driven approaches. They are sensitive but not suited for discovery, as they require prior knowledge about the sample. Like these targeted approaches, SWATH is a DIA method but it does not rely on prior knowledge. The strength of SWATH is that it aims to record all information on the MS^2 level and that this information can be queried post acquisition. However, data interpretation is much more challenging. The theoretical advantage of SWATH that it record all MS features, is in practice limited by mass spectrometric performance. The limited speed and accuracy of the current mass spectrometer generation is not yet sufficient to match the performance of DDA in complex mixtures. In SWATH "useful" information is effectively diluted with signals that cannot be interpreted. In the future, both DDA and DIA could eventually merge, as with very fast MS it would be possible to subject any MS^1 feature or any MS^1 range to fragmentation avoiding the undersampling issue.

Data analysis is a central aspect of proteomics. A recent breakthrough was the development of highly accurate label free algorithms [214]. Especially clinical applications and investigation of primary tissues benefit from the general applicability and almost unlimited scalability of label-free quantification. In combination with the potential to estimate copy numbers, quantitative label-free proteomics has matured into the method of choice for most applications. Nevertheless, labeling strategies such TMT might become more relevant due to their multiplexing capacity. Especially for screens in the clinical context were throughput and accuracy are equally important, the multiplexing capability of TMT has high potential. Another promising but yet unexplored application of MS based proteomics is *de novo* sequencing. *De novo* sequencing faces technical challenges discussed in chapter 2.7. Despite these current limitations, many applications would greatly benefit from *de novo* sequencing. For instance, proteins from diverse or unknown sources like unsequenced microorganisms could be characterized in this way. Finally, fast MS instruments would allow screening for protein modifications or splice variants by identifying the composition of MS^1 features that show interesting patterns in the samples under investigation.

Neurodegeneration is exceptional devastating. It comprises a spectrum of fatal progressing pathologies. While the full awareness of increasing morbidity in ALS is one extreme, progressing loss of one's self in diseases such as Alzheimer's is the other. Theses pathologies are accompanied by immense suffering of patients and kin, underlining the urgent need for therapeutic interventions and therefore a better understanding of the molecular mechanisms. In this thesis I demonstrated how MS based proteomics can be applied to provide novel insights into the molecular mechanisms of neurodegenerative diseases. Furthermore, I presented recent developments in mass spectrometry instrumentation. With the technology advancing, the field of application will expand from basic research to clinical applications and diagnostics and eventually, early diagnosis of neurodegeneration and other diseases might be possible as routine procedure.

In my opinion, the greatest achievement in the history of humankind is the concept of scientific theory. It makes us aware of logical fallacies and superstitious believes, originating from the often misleading intuition that is partially imprinted in our genes. Science provides the most comprehensive understanding of reality and it is the purest, most promising method to learn from the past, understand the present and shape the future.

References

1. Dufour, E. & Larsson, N.G. Understanding aging: revealing order out of chaos. *Biochim Biophys Acta* **1658**, 122-132 (2004).
2. Oeppen, J. & Vaupel, J.W. Demography. Broken limits to life expectancy. *Science* **296**, 1029-1031 (2002).
3. Alper, T., Cramp, W.A., Haig, D.A. & Clarke, M.C. Does the Agent of Scrapie Replicate without Nucleic Acid ? *Nature* **214**, 764-766 (1967).
4. Alper, T., Haig, D.A. & Clarke, M.C. The exceptionally small size of the scrapie agent. *Biochemical and biophysical research communications* **22**, 278-284 (1966).
5. Scheltema, R.A.*, et al.* The Q Exactive HF, a Benchtop Mass Spectrometer with a Pre-filter, High-performance Quadrupole and an Ultra-high-field Orbitrap Analyzer. *Mol Cell Proteomics* **13**, 3698-3708 (2014).
6. Amor, S.*, et al.* Inflammation in neurodegenerative diseases--an update. *Immunology* **142**, 151-166 (2014).
7. Alzheimers, A. 2012 Alzheimer's disease facts and figures. *Alzheimers. Dement.* **8**, 131-168 (2012).
8. Olsen, J.V., Ong, S.E. & Mann, M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics* **3**, 608-614 (2004).
9. Griffith, J.S. Self-replication and scrapie. *Nature* **215**, 1043-1044 (1967).
10. Demartini, D.R. *A Short Overview of the Components in Mass Spectrometry Instrumentation for Proteomics Analyses*, (2013).
11. Newman, E.A. New roles for astrocytes: Regulation of synaptic transmission. *Trends in Neurosciences* **26**, 536-542 (2003).
12. van Deemter, J.J., Zuiderweg, F.J. & Klinkenberg, A. Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography. *Chemical Engineering Science* **50**, 3869-3882 (1995).
13. Makarov, A.*, et al.* Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal Chem* **78**, 2113-2120 (2006).
14. Zwanzig, R., Szabo, A. & Bagchi, B. Levinthal's paradox. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 20-22 (1992).
15. Makarov, A. & Denisov, E. Dynamics of ions of intact proteins in the Orbitrap mass analyzer. *Journal of the American Society for Mass Spectrometry* **20**, 1486-1495 (2009).
16. Uversky, V.N.*, et al.* Pathological unfoldomics of uncontrolled chaos: intrinsically disordered proteins and human diseases. *Chemical reviews* **114**, 6844-6879 (2014).
17. Hartl, F.U., Bracher, A. & Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature* **475**, 324-332 (2011).
18. Tang, Y. & Le, W. Differential Roles of M1 and M2 Microglia in Neurodegenerative Diseases. *Mol Neurobiol* (2015).
19. Prusiner, S.B. Novel proteinaceous infectious particles cause scrapie. *Science* **216**, 136-144 (1982).
20. Gabizon, R., McKinley, M.P., Groth, D. & Prusiner, S.B. Immunoaffinity purification and neutralization of scrapie prion infectivity. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 6617-6621 (1988).
21. Mermet, J.-M., Otto, M. & Valcárcel Cases, M. *Analytical chemistry : a modern approach to analytical science*, (Wiley-VCH, Weinheim ; Great Britain, 2004).
22. Deleault, N.R., Harris, B.T., Rees, J.R. & Supattapone, S. Formation of native prions from minimal components in vitro. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 9741-9746 (2007).
23. Vassar, R. BACE1 inhibitor drugs in clinical trials for Alzheimer's disease. *Alzheimers Res Ther* **6**, 89 (2014).
24. Tyson, N.d. Why You Will Never Find Scientists Leading Armies Into Battle. in *Science*, Vol. 2014 (ed. Kelly, R.) (Richard Dawkins Foundation, Washington, DC 20005, 2014).

25. Kaku, M. Behold the Most Complicated Object in the Known Universe. in *The Leonard Lopate Show* (ed. Lopate, L.) (WNYC Radio, online, 2014).

26. Azevedo, F.A.*, et al.* Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *The Journal of comparative neurology* **513**, 532-541 (2009).

27. Feinstein, S. *From the brain to the classroom : the encyclopedia of learning*, (Greenwood, Santa Barbara, California, 2014).

28. Steinberg, L.D. *Lifespan development : infancy through adulthood*, (Wadsworth Cengage Learning, Belmont, CA, 2011).

29. Squire, L.R. *Fundamental neuroscience*, (Elsevier / Academic Press, Amsterdam ; Boston, 2008).

30. Rodriguez-Arellano, J.J., Parpura, V., Zorec, R. & Verkhratsky, A. Astrocytes in physiological aging and Alzheimer's disease. *Neuroscience* (2015).

31. Cohen, E. & Dillin, A. The insulin paradox: aging, proteotoxicity and neurodegeneration. *Nature reviews. Neuroscience* **9**, 759-767 (2008).

32. Douglas, P.M. & Dillin, A. Protein homeostasis and aging in neurodegeneration. *J Cell Biol* **190**, 719-729 (2010).

33. Kikis, E.A., Gidalevitz, T. & Morimoto, R.I. Protein homeostasis in models of aging and age-related conformational disease. *Advances in experimental medicine and biology* **694**, 138-159 (2010).

34. Irwin, D.J., Lee, V.M. & Trojanowski, J.Q. Parkinson's disease dementia: convergence of alpha-synuclein, tau and amyloid-beta pathologies. *Nature reviews. Neuroscience* **14**, 626-636 (2013).

35. Olzscha, H.*, et al.* Amyloid-like aggregates sequester numerous metastable proteins with essential cellular functions. *Cell* **144**, 67-78 (2011).

36. Jucker, M. & Walker, L.C. Self-propagation of pathogenic protein aggregates in neurodegenerative diseases. *Nature* **501**, 45-51 (2013).

37. Hipp, M.S., Park, S.H. & Hartl, F.U. Proteostasis impairment in protein-misfolding and -aggregation diseases. *Trends in cell biology* **24**, 506-514 (2014).

38. Abel, O.*, et al.* Development of a Smartphone App for a Genetics Website: The Amyotrophic Lateral Sclerosis Online Genetics Database (ALSoD). *JMIR mHealth and uHealth* **1**, e18 (2013).

39. Rademakers, R., Cruts, M. & Van Broeckhoven, C. Genetics of early-onset Alzheimer dementia. *TheScientificWorldJournal* **3**, 497-519 (2003).

40. Bossy-Wetzel, E., Schwarzenbacher, R. & Lipton, S.A. Molecular pathways to neurodegeneration. *Nat Med* **10 Suppl**, S2-9 (2004).

41. Garden, G.A. & La Spada, A.R. Intercellular (mis)communication in neurodegenerative disease. *Neuron* **73**, 886-901 (2012).

42. Obermeier, B., Daneman, R. & Ransohoff, R.M. Development, maintenance and disruption of the blood-brain barrier. *Nat Med* **19**, 1584-1596 (2013).

43. Ramanan, V.K. & Saykin, A.J. Pathways to neurodegeneration: mechanistic insights from GWAS in Alzheimer's disease, Parkinson's disease, and related disorders. *American journal of neurodegenerative disease* **2**, 145-175 (2013).

44. Sulzer, D. Multiple hit hypotheses for dopamine neuron loss in Parkinson's disease. *Trends in Neurosciences* **30**, 244-250 (2007).

45. Rubinsztein, D.C. The roles of intracellular protein-degradation pathways in neurodegeneration. *Nature* **443**, 780-786 (2006).

46. Selkoe, D.J. Folding proteins in fatal ways. *Nature* **426**, 900-904 (2003).

47. Robberecht, W. & Philips, T. The changing scene of amyotrophic lateral sclerosis. *Nature reviews. Neuroscience* **14**, 248-264 (2013).

48. Mori, K.*, et al.* The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science (New York, N.Y.)* **339**, 1335-1338 (2013).

49. Lillo, P. & Hodges, J.R. Frontotemporal dementia and motor neurone disease: overlapping clinic-pathological disorders. *Journal of clinical neuroscience : official journal of the Neurosurgical Society of Australasia* **16**, 1131-1135 (2009).

50. Al-Chalabi, A.*, et al.* An estimate of amyotrophic lateral sclerosis heritability using twin data. *J Neurol Neurosur Ps* **81**, 1324-1326 (2010).

51. DeJesus-Hernandez, M.*, et al.* Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72**, 245-256 (2011).

52. Renton, A.E.*, et al.* A Hexanucleotide Repeat Expansion in C9ORF72 Is the Cause of Chromosome 9p21-Linked ALS-FTD. *Neuron* **72**, 257-268 (2011).

53. Majounie, E.*, et al.* Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. *Lancet neurology* **11**, 323-330 (2012).

54. DeJesus-Hernandez, M.*, et al.* Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS. *Neuron* **72**, 245-256 (2011).

55. Cruts, M.*, et al.* Null mutations in progranulin cause ubiquitin-positive frontotemporal dementia linked to chromosome 17q21. *Nature* **442**, 920-924 (2006).

56. Baker, M.*, et al.* Mutations in progranulin cause tau-negative frontotemporal dementia linked to chromosome 17. *Nature* **442**, 916-919 (2006).

57. Philips, A.V., Timchenko, L.T. & Cooper, T.A. Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science* **280**, 737-741 (1998).

58. Pearson, C.E. Repeat associated non-ATG translation initiation: one DNA, two transcripts, seven reading frames, potentially nine toxic entities! *PLoS Genet* **7**, e1002018 (2011).

59. Ash, Peter E.A.*, et al.* Unconventional Translation of C9ORF72 GGGGCC Expansion Generates Insoluble Polypeptides Specific to c9FTD/ALS. *Neuron* (2013).

60. MDA. Muscular Dystrophy Association. Vol. 2015 (2015).

61. Guerreiro, R., Bras, J. & Hardy, J. SnapShot: Genetics of ALS and FTD. *Cell* **160**, 798-798 e791 (2015).

62. Wojciechowska, M., Olejniczak, M., Galka-Marciniak, P., Jazurek, M. & Krzyzosiak, W.J. RAN translation and frameshifting as translational challenges at simple repeats of human neurodegenerative disorders. *Nucleic acids research* **42**, 11849-11864 (2014).

63. Touriol, C.*, et al.* Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. *Biology of the cell / under the auspices of the European Cell Biology Organization* **95**, 169-178 (2003).

64. Zu, T.*, et al.* Non-ATG-initiated translation directed by microsatellite expansions. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 260-265 (2011).

65. Anfinsen, C.B. Principles that govern the folding of protein chains. *Science* **181**, 223-230 (1973).

66. Dill, K.A. & Chan, H.S. From Levinthal to pathways to funnels. *Nature structural biology* **4**, 10-19 (1997).

67. Wilkinson, B. & Gilbert, H.F. Protein disulfide isomerase. *Biochim Biophys Acta* **1699**, 35-44 (2004).

68. Fischer, G. & Bang, H. The refolding of urea-denatured ribonuclease A is catalyzed by peptidyl-prolyl cis-trans isomerase. *Biochim Biophys Acta* **828**, 39-42 (1985).

69. Jahn, T.R. & Radford, S.E. The Yin and Yang of protein folding. *The FEBS journal* **272**, 5962-5970 (2005).

70. Vendruscolo, M., Paci, E., Karplus, M. & Dobson, C.M. Structures and relative free energies of partially folded states of proteins. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 14817-14821 (2003).

71. Radford, S.E., Dobson, C.M. & Evans, P.A. The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature* **358**, 302-307 (1992).

72. Capaldi, A.P., Shastry, M.C.R., Kleanthous, C., Roder, H. & Radford, S.E. Ultrarapid mixing experiments reveal that Im7 folds via an on-pathway intermediate. *Nat Struct Mol Biol* **8**, 68-72 (2001).

73. Bollen, Y.J., Sanchez, I.E. & van Mierlo, C.P. Formation of on- and off-pathway intermediates in the folding kinetics of Azotobacter vinelandii apoflavodoxin. *Biochemistry* **43**, 10475-10489 (2004).

74. Guijarro, J.I., Sunde, M., Jones, J.A., Campbell, I.D. & Dobson, C.M. Amyloid fibril formation by an SH3 domain. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 4224-4228 (1998).

75. Young, J.C., Agashe, V.R., Siegers, K. & Hartl, F.U. Pathways of chaperone-mediated protein folding in the cytosol. *Nat Rev Mol Cell Biol* **5**, 781-791 (2004).

76. Clark, P.L. Protein folding in the cell: reshaping the folding funnel. *Trends in biochemical sciences* **29**, 527-534 (2004).

77. Barral, J.M., Broadley, S.A., Schaffar, G. & Hartl, F.U. Roles of molecular chaperones in protein misfolding diseases. *Seminars in cell & developmental biology* **15**, 17-29 (2004).

78. Zhang, T.*, et al.* Flexible connection of the N-terminal domain in ClpB modulates substrate binding and the aggregate reactivation efficiency. *Proteins* **80**, 2758-2768 (2012).

79. Wang, G. & Mao, Z. Chaperone-mediated autophagy: roles in neurodegeneration. *Translational neurodegeneration* **3**, 20 (2014).

80.   Aguilaniu, H., Gustafsson, L., Rigoulet, M. & Nyström, T. Asymmetric Inheritance of Oxidatively Damaged Proteins During Cytokinesis. *Science* **299**, 1751-1753 (2003).

81.   Tonoki, A.*, et al.* Genetic evidence linking age-dependent attenuation of the 26S proteasome with the aging process. *Molecular and cellular biology* **29**, 1095-1106 (2009).

82.   Cuervo, A.M. & Dice, J.F. Age-related decline in chaperone-mediated autophagy. *J Biol Chem* **275**, 31505-31513 (2000).

83.   David, D.C.*, et al.* Widespread protein aggregation as an inherent part of aging in C. elegans. *PLoS biology* **8**, e1000450 (2010).

84.   Demontis, F. & Perrimon, N. FOXO/4E-BP signaling in Drosophila muscles regulates organism-wide proteostasis during aging. *Cell* **143**, 813-825 (2010).

85.   Walther, D.M.*, et al.* Widespread Proteome Remodeling and Aggregation in Aging C. elegans. *Cell* **161**, 919-932 (2015).

86.   Reis-Rodrigues, P.*, et al.* Proteomic analysis of age-dependent changes in protein solubility identifies genes that modulate lifespan. *Aging cell* **11**, 120-127 (2012).

87.   Damme, M., Suntio, T., Saftig, P. & Eskelinen, E.L. Autophagy in neuronal cells: general principles and physiological and pathological functions. *Acta neuropathologica* (2014).

88.   Arroyo, D.S.*, et al.* Autophagy in inflammation, infection, neurodegeneration and cancer. *International immunopharmacology* **18**, 55-65 (2013).

89.   Blokhuis, A.M., Groen, E.J., Koppers, M., van den Berg, L.H. & Pasterkamp, R.J. Protein aggregation in amyotrophic lateral sclerosis. *Acta neuropathologica* **125**, 777-794 (2013).

90.   Mackenzie, I.R.*, et al.* Dipeptide repeat protein pathology in C9ORF72 mutation cases: clinico-pathological correlations. *Acta neuropathologica* **126**, 859-879 (2013).

91.   Booth, D.R.*, et al.* Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature* **385**, 787-793 (1997).

92.   Hammarstrom, P., Jiang, X., Hurshman, A.R., Powers, E.T. & Kelly, J.W. Sequence-dependent denaturation energetics: A major determinant in amyloid disease diversity. *Proceedings of the National Academy of Sciences of the United States of America* **99 Suppl 4**, 16427-16432 (2002).

93.   Gupta, R.*, et al.* Firefly luciferase mutants as sensors of proteome stress. *Nature methods* **8**, 879-884 (2011).

94.   Floege, J. & Ehlerding, G. Beta-2-microglobulin-associated amyloidosis. *Nephron* **72**, 9-26 (1996).

95.   Jarrett, J.T. & Lansbury, P.T., Jr. Seeding "one-dimensional crystallization" of amyloid: a pathogenic mechanism in Alzheimer's disease and scrapie? *Cell* **73**, 1055-1058 (1993).

96.   Uversky, V.N. & Fink, A.L. Conformational constraints for amyloid fibrillation: the importance of being unfolded. *Biochim Biophys Acta* **1698**, 131-153 (2004).

97.   Wiedersich, J., Kohler, S., Skerra, A. & Friedrich, J. Temperature and pressure dependence of protein stability: the engineered fluorescein-binding lipocalin FluA shows an elliptic phase diagram. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 5756-5761 (2008).

98.   Tycko, R. Progress towards a molecular-level structural understanding of amyloid fibrils. *Current opinion in structural biology* **14**, 96-103 (2004).

99.   Sunde, M. & Blake, C. The structure of amyloid fibrils by electron microscopy and X-ray diffraction. *Advances in protein chemistry* **50**, 123-159 (1997).

100.   Balch, W.E., Morimoto, R.I., Dillin, A. & Kelly, J.W. Adapting proteostasis for disease intervention. *Science* **319**, 916-919 (2008).

101.   Glabe, C.G. Structural classification of toxic amyloid oligomers. *J Biol Chem* **283**, 29639-29643 (2008).

102.   Bucciantini, M.*, et al.* Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* **416**, 507-511 (2002).

103.   Haass, C. & Selkoe, D.J. Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid beta-peptide. *Nat Rev Mol Cell Biol* **8**, 101-112 (2007).

104.   Morimoto, R.I. Proteotoxic stress and inducible chaperone networks in neurodegenerative disease and aging. *Genes & development* **22**, 1427-1438 (2008).

105.   Last, N.B., Rhoades, E. & Miranker, A.D. Islet amyloid polypeptide demonstrates a persistent capacity to disrupt membrane integrity. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 9460-9465 (2011).

106.   Labbadia, J. & Morimoto, R.I. Huntington's disease: underlying molecular mechanisms and emerging concepts. *Trends in biochemical sciences* **38**, 378-385 (2013).

107. MacDonald, M.E.*, et al.* A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72**, 971-983 (1993).
108. Soto, C. Unfolding the role of protein misfolding in neurodegenerative diseases. *Nature reviews. Neuroscience* **4**, 49-60 (2003).
109. Tourette, C.*, et al.* A large scale Huntingtin protein interaction network implicates Rho GTPase signaling pathways in Huntington disease. *The Journal of biological chemistry* **289**, 6709-6726 (2014).
110. Muchowski, P.J.*, et al.* Hsp70 and hsp40 chaperones can inhibit self-assembly of polyglutamine proteins into amyloid-like fibrils. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 7841-7846 (2000).
111. Behrends, C.*, et al.* Chaperonin TRiC promotes the assembly of polyQ expansion proteins into nontoxic oligomers. *Molecular cell* **23**, 887-897 (2006).
112. Soto, C. & Castilla, J. The controversial protein-only hypothesis of prion propagation. *Nat Med* **10 Suppl**, S63-67 (2004).
113. Collinge, J. Variant Creutzfeldt-Jakob disease. *Lancet* **354**, 317-323 (1999).
114. Will, R.G.*, et al.* A new variant of Creutzfeldt-Jakob disease in the UK. *Lancet* **347**, 921-925 (1996).
115. Collinge, J. Prion diseases of humans and animals: their causes and molecular basis. *Annual review of neuroscience* **24**, 519-550 (2001).
116. Chernoff, Y.O. Mutations and natural selection in the protein world. *Journal of molecular biology* **413**, 525-526 (2011).
117. Uptain, S.M. & Lindquist, S. Prions as protein-based genetic elements. *Annual review of microbiology* **56**, 703-741 (2002).
118. Prusiner, S.B. Prions. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 13363-13383 (1998).
119. Alberti, S., Halfmann, R., King, O., Kapila, A. & Lindquist, S. A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell* **137**, 146-158 (2009).
120. Chernoff, Y.O. Prion: disease or relief? *Nature cell biology* **10**, 1019-1021 (2008).
121. Halfmann, R. & Lindquist, S. Epigenetics in the extreme: prions and the inheritance of environmentally acquired traits. *Science* **330**, 629-632 (2010).
122. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. & Whitehouse, C.M. Electrospray ionization–principles and practice. *Mass spectrometry reviews* **9**, 37-70 (1990).
123. Karas, M. & Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* **60**, 2299-2301 (1988).
124. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. & Whitehouse, C.M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64-71 (1989).
125. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198-207 (2003).
126. Link, A.J.*, et al.* Direct analysis of protein complexes using mass spectrometry. *Nature biotechnology* **17**, 676-682 (1999).
127. Yates, J.R., Ruse, C.I. & Nakorchevsky, A. Proteomics by mass spectrometry: approaches, advances, and applications. *Annual review of biomedical engineering* **11**, 49-79 (2009).
128. Pertea, M. & Salzberg, S.L. Between a chicken and a grape: estimating the number of human genes. *Genome biology* **11**, 206 (2010).
129. Geiger T, W.A., Schaab C, Cox J, Mann M. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol Cell Proteomics* **11**(2012).
130. Khoury, G.A., Baliban, R.C. & Floudas, C.A. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Scientific reports* **1**(2011).
131. Pan, Q., Shai, O., Lee, L.J., Frey, B.J. & Blencowe, B.J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature genetics* **40**, 1413-1415 (2008).
132. Nagaraj, N.*, et al.* Deep proteome and transcriptome mapping of a human cancer cell line. *Molecular systems biology* **7**, 548 (2011).
133. Beck, M.*, et al.* The quantitative proteome of a human cell line. *Molecular systems biology* **7**, 549 (2011).
134. Olsen, J.V., Nielsen, P.A., Andersen, J.R., Mann, M. & Wisniewski, J.R. Quantitative proteomic profiling of membrane proteins from the mouse brain cortex, hippocampus, and cerebellum using the HysTag reagent: mapping of neurotransmitter receptors and ion channels. *Brain research* **1134**, 95-106 (2007).

135. Meissner, F. & Mann, M. Quantitative shotgun proteomics: considerations for a high-quality workflow in immunology. *Nature immunology* **15**, 112-117 (2014).

136. Borner, G.H.*, et al.* Fractionation profiling: a fast and versatile approach for mapping vesicle proteomes and protein-protein interactions. *Molecular biology of the cell* **25**, 3178-3194 (2014).

137. Sharma, K.*, et al.* Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. *Cell reports* **8**, 1583-1594 (2014).

138. Gingras, A.C., Gstaiger, M., Raught, B. & Aebersold, R. Analysis of protein complexes using mass spectrometry. *Nat Rev Mol Cell Biol* **8**, 645-654 (2007).

139. Fields, S. & Song, O. A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245-246 (1989).

140. Ong, S.E. & Mann, M. Mass spectrometry-based proteomics turns quantitative. *Nature chemical biology* **1**, 252-262 (2005).

141. Keilhauer, E.C., Hein, M.Y. & Mann, M. Accurate Protein Complex Retrieval by Affinity Enrichment Mass Spectrometry (AE-MS) Rather than Affinity Purification Mass Spectrometry (AP-MS). *Mol Cell Proteomics* **14**, 120-135 (2015).

142. Wisniewski, J.R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nature methods* **6**, 359-362 (2009).

143. Glatter, T.*, et al.* Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion. *J Proteome Res* **11**, 5145-5156 (2012).

144. Nagaraj, N.*, et al.* Deep proteome and transcriptome mapping of a human cancer cell line. *Molecular systems biology* **7**, 548 (2011).

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

146. Michalski, A., Cox, J. & Mann, M. More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *J Proteome Res* **10**, 1785-1793 (2011).

147. Kocher, T., Swart, R. & Mechtler, K. Ultra-high-pressure RPLC hyphenated to an LTQ-Orbitrap Velos reveals a linear relation between peak capacity and number of identified peptides. *Anal Chem* **83**, 2699-2704 (2011).

148. Jorgenson, J.W. Capillary liquid chromatography at ultrahigh pressures. *Annual review of analytical chemistry* **3**, 129-150 (2010).

149. Smith, R.D., Shen, Y. & Tang, K. Ultrasensitive and quantitative analyses from combined separations-mass spectrometry for the characterization of proteomes. *Accounts of chemical research* **37**, 269-278 (2004).

150. Cristobal, A.*, et al.* In-house construction of a UHPLC system enabling the identification of over 4000 protein groups in a single analysis. *The Analyst* **137**, 3541-3548 (2012).

151. Thakur, S.S.*, et al.* Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. *Mol Cell Proteomics* **10**, M110 003699 (2011).

152. Taylor, G. Disintegration of Water Drops in an Electric Field. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **280**, 383-397 (1964).

153. Mann, M. & Wilm, M. Electrospray mass spectrometry for protein characterization. *Trends in biochemical sciences* **20**, 219-224 (1995).

154. Grimm, R.L. & Beauchamp, J.L. Evaporation and discharge dynamics of highly charged multicomponent droplets generated by electrospray ionization. *The journal of physical chemistry. A* **114**, 1411-1419 (2010).

155. Konermann, L., McAllister, R.G. & Metwally, H. Molecular Dynamics simulations of the electrospray process: formation of NaCl clusters via the charged residue mechanism. *The journal of physical chemistry. B* **118**, 12025-12033 (2014).

156. Scigelova, M., Hornshaw, M., Giannakopulos, A. & Makarov, A. Fourier transform mass spectrometry. *Mol Cell Proteomics* **10**, M111 009431 (2011).

157. Marshall, A.G. & Hendrickson, C.L. High-resolution mass spectrometers. *Annual review of analytical chemistry* **1**, 579-599 (2008).

158. Xian, F., Hendrickson, C.L. & Marshall, A.G. High resolution mass spectrometry. *Anal Chem* **84**, 708-719 (2012).

159. Perry, R.H., Cooks, R.G. & Noll, R.J. Orbitrap mass spectrometry: instrumentation, ion motion and applications. *Mass spectrometry reviews* **27**, 661-699 (2008).
160. Cox, J.*, et al.* Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *Journal of Proteome Research* **10**, 1794-1805 (2011).
161. Domon, B. & Aebersold, R. Options and considerations when selecting a quantitative proteomics strategy. *Nature biotechnology* **28**, 710-721 (2010).
162. ASMS. Vol. 2015 (2013).
163. B.A. Mamyrin, V.I.K., D.V. Shmikk, V.A. Zagulin. The mass-reflectron, a new nonmagnetic time-of-flight mass spectrometer with high resolution. *Journal of Experimental and Theoretical Physics* **37**, 45 (1973).
164. Paul, W. & Steinwedel, H. Ein neues Massenspektrometer ohne Magnetfeld. *Zeitschrift für Naturforschung A* **8**, 448-450 (1953).
165. Paul, W., Reinhard, H.P. & von Zahn, U. Das elektrische Massenfilter als Massenspektrometer und Isotopentrenner. *Z. Physik* **152**, 143-182 (1958).
166. March, R.E. & Todd, J.F.J. *Quadrupole ion trap mass spectrometry*, (J. Wiley, Hoboken, N.J., 2005).
167. Hager, J.W. A new linear ion trap mass spectrometer. *Rapid Communications in Mass Spectrometry* **16**, 512-526 (2002).
168. Schwartz, J.C., Senko, M.W. & Syka, J.E. A two-dimensional quadrupole ion trap mass spectrometer. *Journal of the American Society for Mass Spectrometry* **13**, 659-669 (2002).
169. Mayya, V., Rezaul, K., Cong, Y.S. & Han, D. Systematic comparison of a two-dimensional ion trap and a three-dimensional ion trap mass spectrometer in proteomics. *Mol Cell Proteomics* **4**, 214-223 (2005).
170. Hardman, M. & Makarov, A.A. Interfacing the orbitrap mass analyzer to an electrospray ion source. *Anal Chem* **75**, 1699-1705 (2003).
171. Makarov, A. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal Chem* **72**, 1156-1162 (2000).
172. Kingdon, K.H. A Method for the Neutralization of Electron Space Charge by Positive Ionization at Very Low Gas Pressures. *Physical Review* **21**, 408-418 (1923).
173. Hu, Q.*, et al.* The Orbitrap: a new mass spectrometer. *Journal of mass spectrometry : JMS* **40**, 430-443 (2005).
174. Scigelova, M. & Makarov, A. Orbitrap mass analyzer--overview and applications in proteomics. *Proteomics* **6 Suppl 2**, 16-21 (2006).
175. Michalski, A.*, et al.* Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes. *Molecular & cellular proteomics : MCP* **11**, O111.013698 (2012).
176. Lange, O., Damoc, E., Wieghaus, A. & Makarov, A. Enhanced Fourier transform for Orbitrap mass spectrometry. *International Journal of Mass Spectrometry* **369**, 16-22 (2014).
177. Olsen, J.V.*, et al.* A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Mol Cell Proteomics* **8**, 2759-2769 (2009).
178. Mann, M. & Kelleher, N.L. Precision proteomics: the case for high resolution and high mass accuracy. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 18132-18138 (2008).
179. Michalski, A., Neuhauser, N., Cox, J. & Mann, M. A systematic investigation into the nature of tryptic HCD spectra. *J Proteome Res* **11**, 5479-5491 (2012).
180. Olsen, J.V.*, et al.* Higher-energy C-trap dissociation for peptide modification analysis. *Nature methods* **4**, 709-712 (2007).
181. Wells, J.M. & McLuckey, S.A. Collision-induced dissociation (CID) of peptides and proteins. *Methods in enzymology* **402**, 148-185 (2005).
182. Steen, H. & Mann, M. The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol* **5**, 699-711 (2004).
183. McAlister, G.C., Phanstiel, D.H., Brumbaugh, J., Westphall, M.S. & Coon, J.J. Higher-energy collision-activated dissociation without a dedicated collision cell. *Mol Cell Proteomics* **10**, O111 009456 (2011).
184. Zubarev, R.A.*, et al.* Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal Chem* **72**, 563-573 (2000).

185. Syka, J.E., Coon, J.J., Schroeder, M.J., Shabanowitz, J. & Hunt, D.F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 9528-9533 (2004).

186. Mikesh, L.M.*, et al.* The utility of ETD mass spectrometry in proteomic analysis. *Biochim Biophys Acta* **1764**, 1811-1822 (2006).

187. Frese, C.K.*, et al.* Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal Chem* **84**, 9668-9673 (2012).

188. Michalski A, D.E., Hauschild JP, Lange O, Wieghaus A, Makarov A, Nagaraj N, Cox J, Mann M, Horning S. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol Cell Proteomics* **10**(2011).

189. Second, T.P.*, et al.* Dual-pressure linear ion trap mass spectrometer improving the analysis of complex protein mixtures. *Anal Chem* **81**, 7757-7765 (2009).

190. Makarov, A., Denisov, E., Lange, O. & Horning, S. Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer. *Journal of the American Society for Mass Spectrometry* **17**, 977-982 (2006).

191. Gillet, L.C.*, et al.* Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics* **11**, O111 016717 (2012).

192. Venable, J.D., Dong, M.Q., Wohlschlegel, J., Dillin, A. & Yates, J.R. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nature methods* **1**, 39-45 (2004).

193. Selevsek, N.*, et al.* Reproducible and consistent quantification of the Saccharomyces cerevisiae proteome by SWATH-MS. *Mol Cell Proteomics* (2015).

194. Peterson, A.C., Russell, J.D., Bailey, D.J., Westphall, M.S. & Coon, J.J. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol Cell Proteomics* **11**, 1475-1488 (2012).

195. Perkins, D.N., Pappin, D.J., Creasy, D.M. & Cottrell, J.S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567 (1999).

196. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* **26**, 1367-1372 (2008).

197. Cox, J.*, et al.* A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. *Nat Protoc* **4**, 698-705 (2009).

198. Elias, J.E. & Gygi, S.P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature methods* **4**, 207-214 (2007).

199. Wilhelm, M.*, et al.* Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582-587 (2014).

200. Kim, M.S.*, et al.* A draft map of the human proteome. *Nature* **509**, 575-581 (2014).

201. Ezkurdia, I., Vazquez, J., Valencia, A. & Tress, M. Analyzing the First Drafts of the Human Proteome. *J Proteome Res* (2014).

202. Bantscheff, M., Lemeer, S., Savitski, M.M. & Kuster, B. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Analytical and bioanalytical chemistry* **404**, 939-965 (2012).

203. Ong, S.E.*, et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**, 376-386 (2002).

204. Schwanhausser, B., Gossen, M., Dittmar, G. & Selbach, M. Global analysis of cellular protein translation by pulsed SILAC. *Proteomics* **9**, 205-209 (2009).

205. Kruger, M.*, et al.* SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. *Cell* **134**, 353-364 (2008).

206. Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J.R. & Mann, M. Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nature methods* **7**, 383-385 (2010).

207. Hsu, J.L., Huang, S.Y., Chow, N.H. & Chen, S.H. Stable-isotope dimethyl labeling for quantitative proteomics. *Anal Chem* **75**, 6843-6852 (2003).

208. Gygi, S.P.*, et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature biotechnology* **17**, 994-999 (1999).

209. Thompson, A.*, et al.* Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **75**, 1895-1904 (2003).

210. Ross, P.L.*, et al.* Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **3**, 1154-1169 (2004).

211. Liu, H., Sadygov, R.G. & Yates, J.R., 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* **76**, 4193-4201 (2004).

212. Ishihama, Y.*, et al.* Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* **4**, 1265-1272 (2005).

213. Lu, P., Vogel, C., Wang, R., Yao, X. & Marcotte, E.M. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nature biotechnology* **25**, 117-124 (2007).

214. Cox, J.*, et al.* MaxLFQ allows accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction. *Mol Cell Proteomics* (2014).

215. Uhlen, M.*, et al.* A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol Cell Proteomics* **4**, 1920-1932 (2005).

216. Zeiler, M., Straube, W.L., Lundberg, E., Uhlen, M. & Mann, M. A Protein Epitope Signature Tag (PrEST) library allows SILAC-based absolute quantification and multiplexed determination of protein copy numbers in cell lines. *Molecular & cellular proteomics : MCP* **11**, O111.009613 (2012).

217. Beynon, R.J., Doherty, M.K., Pratt, J.M. & Gaskell, S.J. Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nature methods* **2**, 587-589 (2005).

218. Kirkpatrick, D.S., Gerber, S.A. & Gygi, S.P. The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. *Methods* **35**, 265-273 (2005).

219. Brownridge, P.*, et al.* Global absolute quantification of a proteome: Challenges in the deployment of a QconCAT strategy. *Proteomics* **11**, 2957-2970 (2011).

220. Schwanhausser, B.*, et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337-342 (2011).

221. Silva, J.C., Gorenstein, M.V., Li, G.Z., Vissers, J.P. & Geromanos, S.J. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol Cell Proteomics* **5**, 144-156 (2006).

222. Wisniewski, J.R.*, et al.* Extensive quantitative remodeling of the proteome between normal colon tissue and adenocarcinoma. *Molecular systems biology* **8**, 611 (2012).

223. Cox, J. & Mann, M. Computational principles of determining and improving mass precision and accuracy for proteome measurements in an Orbitrap. *Journal of the American Society for Mass Spectrometry* **20**, 1477-1485 (2009).

224. Hibbert, D.B. Experimental design in chromatography: a tutorial review. *J Chromatogr B Analyt Technol Biomed Life Sci* **910**, 2-13 (2012).

225. Wiese, S.*, et al.* Isolation and enrichment of embryonic mouse motoneurons from the lumbar spinal cord of individual mouse embryos. *Nat Protoc* **5**, 31-38 (2010).

226. Hornburg, D.*, et al.* Deep proteomic evaluation of primary and cell line motoneuron disease models delineates major differences in neuronal characteristics. *Mol Cell Proteomics* **13**, 3410-3420 (2014).

227. May, S.*, et al.* C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. *Acta neuropathologica* **128**, 485-503 (2014).

228. Ripaud, L.*, et al.* Overexpression of Q-rich prion-like proteins suppresses polyQ cytotoxicity and alters the polyQ interactome. *Proceedings of the National Academy of Sciences of the United States of America* (2014).

229. Compagni, A. & Christofori, G. Recent advances in research on multistage tumorigenesis. *Br J Cancer* **83**, 1-5 (2000).

230. Soldner F Fau - Laganiere, J.*, et al.* Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations.

231. Taylor, A.M.*, et al.* A microfluidic culture platform for CNS axonal injury, regeneration and transport. *Nature methods* **2**, 599-605 (2005).

232. Al-Sarraj, S.*, et al.* p62 positive, TDP-43 negative, neuronal cytoplasmic and intranuclear inclusions in the cerebellum and hippocampus define the pathology of C9orf72-linked FTLD and MND/ALS. *Acta neuropathologica* **122**, 691-702 (2011).

233. Smethurst, P., Sidle, K.C. & Hardy, J. Prion-like mechanisms of TDP-43 in ALS. *Neuropathology and applied neurobiology* (2014).

234. Holmes, B.B. & Diamond, M.I. Cellular mechanisms of protein aggregate propagation. *Current opinion in neurology* **25**, 721-726 (2012).

235. Gonzalez-Porta, M., Frankish, A., Rung, J., Harrow, J. & Brazma, A. Transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene. *Genome biology* **14**, R70 (2013).

236. Lee, Y.*, et al.* Oligodendroglia metabolically support axons and contribute to neurodegeneration. *Nature* **487**, 443-448 (2012).

237. Fernandez, C., Santos, H.M., Ruiz-Romero, C., Blanco, F.J. & Capelo-Martinez, J.L. A comparison of depletion versus equalization for reducing high-abundance proteins in human serum. *Electrophoresis* **32**, 2966-2974 (2011).

238. Kulak, N.A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nature methods* **11**, 319-324 (2014).

239. Hughes, C.S.*, et al.* Ultrasensitive proteome analysis using paramagnetic bead technology. *Molecular systems biology* **10**, 757 (2014).

240. Manadas, B., English, J.A., Wynne, K.J., Cotter, D.R. & Dunn, M.J. Comparative analysis of OFFGel, strong cation exchange with pH gradient, and RP at high pH for first-dimensional separation of peptides from a membrane-enriched protein fraction. *Proteomics* **9**, 5194-5198 (2009).

241. Wisniewski, J.R., Zougman, A. & Mann, M. Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *J Proteome Res* **8**, 5674-5678 (2009).

242. Wang, Y.*, et al.* Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *Proteomics* **11**, 2019-2026 (2011).

243. Horth, P., Miller, C.A., Preckel, T. & Wenz, C. Efficient fractionation and improved protein identification by peptide OFFGEL electrophoresis. *Mol Cell Proteomics* **5**, 1968-1974 (2006).

244. Kanu, A.B., Dwivedi, P., Tam, M., Matz, L. & Hill, H.H., Jr. Ion mobility-mass spectrometry. *Journal of mass spectrometry : JMS* **43**, 1-22 (2008).

245. Valentine, S.J.*, et al.* Toward plasma proteome profiling with ion mobility-mass spectrometry. *J Proteome Res* **5**, 2977-2984 (2006).

Abbreviations

AD	Alzheimer's disease
ALS	Amyotrophic Lateral Sclerosis
BSE	bovine spongiform encephalopathy
CNS	central nervous system
CRM	charge residue model
CID	collision-induced dissociation
CTF-h	C-terminal fragments of APP
DIA	data independent acquisitions
DDA	data dependent acquisition
DoE	design of experiment
DC	direct current
ECD	electron capture dissociation
ETD	electron transfer dissociation
ESI	electrospray ionization
eFT	enhanced Fourier transformation
XIC	extracted ion current
FDR	false discovery rate
FTICR	Fourier transform ion cyclotrons
FT	Fourier transformation
FWHM	full width half maximum
HCD	higher energy c-trap dissociation
HD	Huntington's disease
QTOF	Hybrid quadrupole time-of-flight
iPSC	induced pluripotent stem cells
IEM	ion evaporation model
IMS	ion mobility separation
LTQ	linear ion traps
LTP	long-term potentiation
MS	mass spectrometer
MALDI	Matrix-assisted laser desorption / ionization
NAC	nascent polypeptide associated complex
ND	neurodegenerative disorder
103Q-GFP	N-terminal fragments of huntingtin containing a polyQ stretch
PD	Parkinson's disease
ppb	parts per billion
ppm	parts per million
pDRP	poly dipeptide repeat
polyQ	poly-glutamine
PEP	posterior error probability
PTMs	posttranslational modifications
PrP^C	prion Proteins correctly folded
PrP^{SC}	prion Proteins misfolded

PrESTs	Protein Epitope Signature Tags
RF	radio frequency oscillating alternating current
RAN translation	repeat associated non-ATG translation
RAC	ribosome-associated complex
S / MRM	single or multi reaction monitoring
sAPP-h	soluble APP ectodomain
SMA	Spinobulbar Muscle Atrophy
SILAC	stable amino acid labeling in cell culture
TOF	time of flight
TSE	transmissible spongiform encephalopathies
TF	trigger factor
UPS	ubiquitin / proteasome system

Acknowledgment

Curriculum Vitae

Personal details

Name	Daniel Hornburg
Date of birth	1984/06/05
Place of birth	Magdeburg, Germany
Nationality	German

Education

Since 05/2011	PhD candidate, Proteomics and Signal Transduction (Matthias Mann), Max Planck Institute of Biochemistry
09/2008 – 10/2010	Molecular Biotechnology (M. Sc.), Technical University of Munich (Dieter Langosch), Thesis: *'Investigation of the structural flexibility of APP transmembrane domain and its mutant Aβ26-55 G33I with molecular biological methods and molecular dynamics.'*
08/2008 – 01/2009	Graduate exchange student, Georgia Institute of Technology, USA
10/2005 – 09/2008	Molecular Biotechnology (B. Sc.), Technical University of Munich (Angelika Schnieke), Thesis: *'Construction of TP53 gene targeting vector: Generation of pig with genetic predisposition to pancreatic cancer'*

Conferences

2014	**EUROMOTOR** (Dublin, Ireland) oral presentation; **HUPO** (Madrid, Spain) poster presentation; **ToPAG** (Murnau, Germany) oral presentation
2013	**EUROMOTOR** (Amsterdam, Netherland) oral presentation; 13th **Munich Science Days** (Munich, Germany) poster presentation, Life Science **PhD Symposium** (Munich, Germany) poster presentation
2012	**EUROMOTOR** (Sheffield, England) oral presentation;
2011	**EUROMOTOR** (Leuven, Belgium) oral presentation; **ALSMND** (Chicago, USA) poster presentation

External Trainings

2014	Design of Experiments – Quality by Design and Design Space (Munich, Germany)
2012	Aekta Chromatography (Munich, Germany); "Innovationsakademie Biotechnologie" entrepreneurial training (Berlin, Germany)

Teaching

2014	**"Omics data analysis" Workshop** (Bonn, Germany); one week **practical course** (LMU, Germany)
2013	**MaxQuant Summer School** (Munich, Germany); one week **practical course** (LMU, Germany)
2012	**MaxQuant Summer School** (Munich, Germany); one week **practical course** (LMU, Germany)
2011	**MaxQuant Summer School** (Munich, Germany)
2006 – 2014	Tutor / supervision of practical courses in physics and biochemistry (TUM & LMU, Germany)

Fellowships and Awards

Since 05/2011	International Max Planck Research School for Molecular and Cellular Life Sciences: From Biology to Medicine (IMPRS-LS) fellow
Since 2012	e-fellows fellowship
07/2015	Best Talk Award (First Prize): IMPRS-LS Evaluation 2015
07/2014	Best Talk Award (Second Prize): IMPRS-LS Seminar 2014
03/2010 – 03/2011	EXIST entrepreneur fellowship (BMBF)
08/2008 – 01/2009	TUMexchange (Technical University of Munich) Georgia Institute of Technology, USA

Scientific Publications and Patents (*equal contribution)

Willem M.; Tahirovic S.; Busche M.; Ovsepian S.; Chafai M.; Kootar S.; **Hornburg D.**; Evans L.; et al. (2015): η-Secretase processing of APP inhibits hippocampal neuronal activity. In **Nature,** DOI: 10.1038/nature14864.

Beck S.; Michalski A.; Raether O.; Lubeck M.; Kaspar S.; Goedecke N.; Baessmann C.; **Hornburg D.**; Meier F.; Paron I.; Kulak NA.; Cox J.; Mann M. (2015): The Impact II, a Very High-Resolution Quadrupole Time-of-Flight Instrument (QTOF) for Deep Shotgun Proteomics. In **Mol. Cell Proteomics** 14 (7), pp. 3410–3420. DOI: 10.1074/mcp.M114.047407.

Hornburg D; Drepper C.; Butter F.; Meissner F.; Sendtner M.; Mann M. (2014): Deep proteomic evaluation of primary and cell line motoneuron disease models delineates major differences in neuronal characteristics. In **Mol. Cell Proteomics** 13 (12), pp. 3410–3420. DOI: 10.1074/mcp.M113.037291.

Hornburg D.*; May S.*; Schludi M.*; Arzberger T.; Rentzsch K.; Schwenk B. et al. (2014): C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. In **Acta neuropathologica** 128 (4), pp. 485–503. DOI: 10.1007/s00401-014-1329-4.

Ripaud L.; Chumakova V.; Antonin M.; Hastie A.; Pinkert S.; Körner R.; Ruff K.; Pappu R.; **Hornburg D.**; Mann M.; Hartl U.; Hipp M. (2014): Overexpression of Q-rich prion-like proteins suppresses polyQ cytotoxicity and alters the polyQ interactome. In **PNAS** of the United States of America 111 (51), pp. 18219–18224. DOI: 10.1073/pnas.1421313111.

Scharnagl, Christina; Pester, Oxana; Hornburg, Philipp; **Hornburg, Daniel**; Götz, Alexander; Langosch, Dieter (2014): Side-chain to main-chain hydrogen bonding controls the intrinsic backbone dynamics of the amyloid precursor protein transmembrane helix. In **Biophys. J.** 106 (6), pp. 1318–1326. DOI: 10.1016/j.bpj.2014.02.013.

Scheltema R.; Hauschild J.; Lange O.; **Hornburg D.**; Denisov E.; Damoc E. et al. (2014): The Q Exactive HF, a Benchtop mass spectrometer with a pre-filter, high-performance quadrupole and an ultra-high-field Orbitrap analyzer. In **Mol. Cell Proteomics** 13 (12), pp. 3698–3708. DOI: 10.1074/mcp.M114.043489.

Pester O.; Barrett P.; **Hornburg D.**; Hornburg P.; Pröbstle R.; Widmaier S. et al. (2013): The backbone dynamics of the amyloid precursor protein transmembrane helix provides a rationale for the sequential cleavage mechanism of γ-secretase. In **Journal of the American Chemical Society** 135 (4), pp. 1317–1329. DOI: 10.1021/ja3112093.

Pester O, **Hornburg D.**; Duerrbaum M.; Hornburg P.; Scharnagl C.; Multhaup G.; Langosch D. (2012), Mapping the Backbone Dynamics of the Alzheimer's APP Transmembrane Domain. In **Biophys. J.** 99 (3), pp. 266. DOI: http://dx.doi.org/10.1016/j.bpj.2011.11.1466

Schwarz WH.; Krauss J.; Zverlov VV.; **Hornburg D**.; Köck D.; Schulte LP. (2011), Artificial cellulosome and the use of the same for enzymatic breakdown of resilient substrates, **US Patent** App. 13/810,920

Quint S.; Widmaier S., Minde D.; **Hornburg D.**; Langosch D.; Scharnagl C. (2010): Residue-specific side-chain packing determines the backbone dynamics of transmembrane model helices. In **Biophys. J.** 99 (8), pp. 2541–2549. DOI: 10.1016/j.bpj.2010.08.031.

Munich, 22.05.2015

Daniel Hornburg