The triple power of D³: Protein intrinsic disorder in degenerative diseases

Vladimir N. Uversky¹,²

¹Department of Molecular Medicine, USF Health Byrd Alzheimer’s Research Institute, Morsani College of Medicine, University of South Florida, Tampa, Florida 33612, USA, ²Institute for Biological Instrumentation, Russian Academy of Sciences, 142290 Pushchino, Moscow Region, Russia

TABLE OF CONTENTS

1. Abstract
2. Introduction
   2.1. Neurodegenerative diseases as proteinopathies
   2.2. Neurodegenerative diseases as amyloidoses
   2.3. Prion-like propagation of neurodegenerative diseases
   2.4. Relationships between amino acid sequence and fibrillogenesis
   2.5. Blocking fibrillogenesis
   2.6. Intrinsically disordered proteins, IDPs
      2.6.1. Concept
      2.6.2. Experimental techniques for IDP detection
      2.6.3. Sequence peculiarities of IDPs and predictors of intrinsic disorder
      2.6.4. Natural abundance of IDPs and their biological functions
3. The first D³: Abundance of IDPs in neurodegenerative diseases as evident from the bioinformatics analyses
4. α-Synuclein as an illustrative example of IDPs related to the neurodegeneration
   4.1. Structural properties of α-synuclein: Protein-chameleon
   4.2. α-Synuclein maintains disordered structure in crowded environment
   4.3. α-Synuclein maintains disordered structure side the living cell
   4.4. α-Synuclein as a disordered hub
   4.5. A place of the helical α-synuclein tetramers within a complex structural image of protein-chameleon
5. Neurodegeneration-associated intrinsically disordered proteins and corresponding maladies
   5.1. Amyloid β-protein and Alzheimer’s disease
   5.2. Tau protein in Alzheimer’s disease and other tauopathies
   5.3. Prion protein and prion diseases
   5.4. α-Synuclein and synucleinopathies
      5.4.1. α-Synuclein and Parkinson’s disease
      5.4.2. α-Synuclein in dementia with Levy bodies and other Lewy body disorders
         5.4.2.1. Dementia with Levy Bodies.
         5.4.2.3. Amyotrophic lateral sclerosis-parkinsonism/dementia complex of Guam.
      5.4.2.3. Other Lewy body diseases.
      5.4.3. α-Synuclein and Alzheimer’s disease
      5.4.4. α-Synuclein and Down’s syndrome
      5.4.5. α-Synuclein and Multiple System Atrophy
      5.4.6. α-Synuclein and Neurodegeneration with Brain Iron Accumulation Type 1 (NBIA1)
   5.5. β- and γ-Synucleins in Parkinson’s disease and dementia with Levy bodies
   5.6. Polyglutamine repeat diseases associated with aggregation of huntingtin, ataxins, androgen receptor, and atrophin-1
      5.6.1. Polyglutamine repeat diseases
      5.6.2. Huntingtin and structure of polyglutamine stretches
      5.6.3. Dentatorubral-pallidolysian atrophy protein (atrophin-1)
      5.6.4. Androgen receptor
      5.6.5. Ataxin-2
      5.6.6. Ataxin-3
      5.6.7. P/Q-type calcium channel α1a subunit (CACNA1A)
      5.6.8. Ataxin-7
      5.6.9. TATA-box-binding protein
   5.7. Aβ1 peptide and familial British dementia
   5.8. ADan in familial Danish dementia
   5.9. Gial fibrillary acidic protein and Alexander disease
   5.10. Mitochondrial DNA polymerase γ and Alpers disease
   5.11. DNA excision repair protein ERCC-6 and Cockayne syndrome
   5.12. Survival of motor neurons protein and spinal muscular atrophy
   5.13. Septins in neuropathology

181
Protein intrinsic disorder in degenerative diseases

5.14 Neurotrophin NGF and neurodegeneration
5.15 FUS and TDP-43 in amyotrophic lateral sclerosis and frontotemporal lobar degeneration
5.16 HIV-associated dementia and aggregation of Fas
5.17 Valosin-containing protein in neurodegeneration
5.18 Acetyltransferase p300 in Parkinson’s disease
5.19 Optineurin in neurodegeneration
6. The second D3: Dysfunction of neurodegeneration-promoting IDPs
7. The third D3: IDPs controlling and regulating the neurodegeneration-promoting IDPs
7.1 Disordered chaperones in neurodegeneration: Malleable guardians
7.2 Sirtuins in neurodegenerative diseases
7.3 Stress granules and neurodegeneration
7.4 Tubulin polymerization promoting protein and α-synuclein aggregation
8. Conclusions
9. Acknowledgements
10. References

1. ABSTRACT

This review is an update of an article published four years ago (Uversky V.N. (2009) Intrinsically disordered proteins in neurodegenerative diseases: another illustration of the D3 concept. Frontiers in Bioscience 14, 5188-5238). The major goal of this review is to show the interconnections between intrinsically disordered proteins (IDPs) and human neurodegeneration. This brings to existence a new D3 concept: protein intrinsic Disorder in neuroDegenerative Diseases. An important aspect of the D3 concept is that it deals with three D3s, emphasizing that intrinsically Disordered proteins are abundantly found in various neuroDegenerative Diseases (the first D3), that these IDPs provoke neuroDegeneration due to their Dysfunctionality (the second D3), and that neuroDegeneration-related IDPs are often controlled by other Disordered proteins (the third D3).

2. INTRODUCTION

2.1. Neurodegenerative diseases as proteinopathies

The large class of human neurodegenerative disorders includes many acquired neurological diseases with distinct phenotypic and pathologic expressions, all characterized by the pathological conditions in which cells of the brain and spinal cord are lost. The name for these diseases is derived from a Greek word νευρό-, "nerval" and a Latin verb dēgenerāre, "to decline" or "to worsen". As neurons are not readily regenerated, their deterioration leads over time to dysfunction and disabilities. Neurodegenerative diseases can be divided into two groups according to their phenotypic effects: (i) Conditions causing problems with movements; and (ii) Conditions affecting memory and leading to dementia. Neurodegeneration is a slow process, which begins long before the patient experiences any symptoms. It can take months or even years before visible outcomes of this degeneration are felt and diagnosed. Symptoms are usually noticed when many cells die or fail to function and a part of the brain begins to cease functioning properly. For example, the symptoms of Parkinson’s disease (PD) become apparent after more than ~70% dopaminergic neurons die in substantia nigra (a small area of cells in the mid-brain affected by PD). Similarly, when symptoms of Alzheimer’s disease (AD) start to be visible, significant shrinkage of brain tissue (the most affected areas being cortex and hippocampus) already took place due to the extensive cell death.

Until recently, a link between AD, prion diseases, PD, Huntington’s disease (HD), and several other neurodegenerative disorders was elusive. However, advances in molecular biology, immunopathology and genetics indicated that these diseases might share a common pathophysiology, where disarrangement of a specific protein processing, functioning, and/or folding takes place. Therefore, neurodegenerative disorders represent a set of proteinopathies, which can be classified and grouped based on the causative proteins. In fact, from this viewpoint neurodegenerative disorders represent a subset of a broader class of human diseases known as protein conformational or protein misfolding diseases. These disorders arise from the failure of a specific peptide or protein to adopt its native functional conformational state. The obvious consequences of misfolding are protein aggregation (and/or fibril formation), loss of function, and gain of toxic function. Some proteins have an intrinsic propensity to assume a pathologic conformation, which becomes evident with aging or at persistently high concentrations. It is now believed that the ability of polypeptide chains to form amyloid structures is not restricted to the relatively small number of proteins associated with recognized clinical disorders, but represents a generic property of a polypeptide chain (1). Intriguingly, even such proteins as green fluorescent protein, GFP (2), and molecular chaperonin GroES (3) can be induced to aggregate.

Interactions (or impaired interactions) with some endogenous factors (e.g., chaperones, intracellular or extracellular matrixes, other proteins, small molecules) can change conformation of a pathogenic protein and increase its propensity to misfold. Misfolding can originate from point mutation(s) or result from an exposure to internal or external toxins, impaired posttranslational modifications (phosphorylation, advanced glycation, deamidation, racemization, etc.), an increased probability of degradation, impaired trafficking, lost binding partners or oxidative damage. All these factors can act independently or in association with one another. Table 1 lists some of the IDPs involved in various neurodegenerative diseases. As the
Protein intrinsic disorder in degenerative diseases

The major focus of this review is the neurodegenerative mechanisms of IDPs, the last section of Introduction is devoted to the brief introduction of the protein intrinsic disorder phenomenon.

2.2. Neurodegenerative diseases as amyloidoses

Many of the diseases listed in Table 1 are in fact protein deposition diseases associated with the formation of extracellular amyloid fibrils or intracellular inclusions with amyloid-like characteristics. Protein deposition diseases can be sporadic (85%), hereditary (10%) or even transmissible, as in the case of prion diseases (5%) (4). Although these diseases, being very different clinically, they share similar molecular mechanisms where a specific protein or protein fragment changes from its natural soluble form into insoluble fibrils. It has been pointed out that prior to fibrillation, amyloidogenic polypeptides may be rich in β-sheet, α-helix, or contain both α-helices and β-sheets. They may be globular proteins with rigid 3D-structure or belong to the class of natively unfolded (or intrinsically unstructured) proteins (5). Molecular mechanisms of fibrillation of IDPs and ordered proteins are different (5): in ordered proteins, the first critical step in fibrillogenesis is the partial unfolding (5-15), whereas the earliest stage of fibrillation of IDPs is their partial folding (5). Intriguingly, the fibrillation of ordered proteins is frequently associated with the pathogenesis of systemic amyloidoses (e.g., mutated lysozyme in hereditary systemic amyloidosis, mutated gelsolin in Finish-type familial amyloidosis, β₂-microglobulin in amyloidosis associated with hemodialysis, immunoglobulin light chain variable domains in light chain associated amyloidosis and light chain deposition disease, etc.), but these proteins are not too common in neurodegenerative diseases (5). Cystatin C represents an illustrative example of ordered proteins, fibrillation of which is indirectly related to neurodegeneration. In fact, in hereditary cystatin C amyloid angiopathy, mutated cystatin C forms amyloid, predominantly in brain arteries and arterioles. This amyloid deposition in the vessel walls causes thickening of the walls leading to occlusion or rupture and resulting in brain hemorrhage (16).

Despite significant structural differences of fibrilating proteins, the fibrils from different pathologies display many common properties, including a core cross-β-sheet structure in which continuous β-sheets are formed with β-strands running perpendicular to the long axis of the fibrils (17, 18). This β-pleated sheet structure of fibrils constitutes the basis of the unusual resistance of all kinds of amyloid to degradation and, therefore, the progressive deposition of the material (19). Furthermore, all fibrils have similar twisted, rope-like structures that are typically 7–13 nm wide (17, 20, 21) and consist of a number of protofilaments (typically 2–6), each about 2–5 nm in diameter (20, 21). Alternatively, protofilaments may associate laterally to form long ribbons that are 2–5 nm thick and up to 30 nm wide (22-24).

2.3. Prion-like propagation of neurodegenerative diseases

An intriguing recent development in the field of neurodegenerative diseases is the recognition of the fact that pathogenesis of these diseases might resemble prionopathies (25-29). What meant here is that many features of these diseases (such as phenotypic diversity and the peculiarities of the pathology propagation where aggregates of pathogenic proteins spread within the central nervous system according to a very predictable pattern (30) and where protein aggregates move between cells and seed the misfolding of their normal conformers in recipient neurons in a prion-like manner (29)) are clearly similar to the molecular mechanisms that underlie prion pathogenesis (26, 29). The prion-like disease propagation is externally non-infectious, as in contrast to typical prionopathies, these maladies cannot be transmitted from one person to another (31, 32). However, internally, they are clearly “infectious” since there is a prion-like transmission of protein aggregates between the neurons within the affected brain, with protein aggregates being able to self-propagate by moving from affected to the originally non-affected neurons and by seeding the protein aggregation at points of their arrival.

Although the prion-like propagation was described for many neurodegeneration-related proteins, such as α-synuclein (30, 33-41) in PD, Aβ (42-47) and tau protein (47-53) in AD, huntingtin (27, 54) in HD, and SOD1, FUS, and TDP-43 in amyotrophic lateral sclerosis (ALS) (55-57), the “infectivity” of α-synuclein aggregates is considered below as an illustrative example of this interesting and important phenomenon. The reason for choosing α-synuclein as an illustration of pathological propagation of aggregates is manifold. First, the compelling evidence has been derived from the analysis of patients whose lost nigral dopaminergic neurons were replaced by grafting embryonic neural tissue grafted into the striatum. Unexpectedly, based on the autopsy of deceased patients who were subject to stereotaxical neuron transplantation a decade prior to death, it has been concluded that these transplanted neurons contained α-synuclein pathology (58-62), indicating that aggregated α-synuclein can be transferred directly from the host brain to grafted cells (63). Second, the mechanisms of the intercellular transfer of α-synuclein aggregates are rather well-studied using various models. The consensus model is presented in Figure 1 (37), which shows that misfolded/aggregated α-synuclein can find its way from the sick neurons to the extracellular space (this is achieved via the active release of normal, misfolded, and aggregated α-synuclein by the affected neuron or after the cell death) and from extracellular space to the originally healthy neuron. Once inside a new neuron, misfolded/aggregated α-synuclein starts to serve as an active template that seeds the aggregation of numerous α-synuclein monomers and thereby promotes the formation of the specific proteinaceous hallmarks of PD, Lewy bodies (LBs) or Lewy neurites (LNs) (64).

These recent findings question the use of transplants as potential therapies of neurodegenerative diseases, since such transplanters will be inevitably “infected” by the existing pathogenic aggregates. However, they definitely give new hope since they indicate that some means can be found that would prevent cell-to-cell
## Table 1. IDPs and associated neurodegenerative diseases

<table>
<thead>
<tr>
<th>Protein (number of residues)</th>
<th>Disease(s)</th>
<th>Disorder by prediction (%)</th>
<th>Disorder by experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (42)</td>
<td>Alzheimer’s disease with amloidosis Congophilic angiopathy</td>
<td>16.9 (28.6)</td>
<td>NMR and far-UV CD analyses revealed that the monomeric peptide is highly unfolded</td>
</tr>
<tr>
<td>Tau (758)</td>
<td>Tauopathies</td>
<td>77.6 (99.1)</td>
<td>Tau protein was shown to be in a random coil-like conformation according to far-UV CD, FTIR, X-ray scattering and biochemical assays</td>
</tr>
<tr>
<td>Prion protein (231)</td>
<td>Prion diseases Creutzfeld-Jacob disease Gerstmann-Sträussler-Scheinker syndrome Fatal familial insomnia Kuru Bovine spongiform encephalopathy Scrapie Chronic wasting disease</td>
<td>55.8 (61.0)</td>
<td>According to NMR and far-UV CD, the N-terminal region (from amino acid 23 to 126) is largely unstructured in the isolated molecule in solution</td>
</tr>
<tr>
<td>α-Synuclein (140)</td>
<td>Synucleinopathies Parkinson’s disease Lewy body variant of Alzheimer’s disease Diffuse Lewy body disease Dementia with Lewy bodies Multiple system atrophy Neurodegeneration with brain iron accumulation type I</td>
<td>90.7 (37.1)</td>
<td>Highly unfolded structure of entire protein is confirmed by NMR, FTIR, SAXS, far-UV CD, gel filtration, dynamic light scattering, FRET, limited proteolysis, aberrant mobility in SDS-PAGE</td>
</tr>
<tr>
<td>β-Synuclein (134)</td>
<td>Parkinson’s disease Diffuse Lewy body disease</td>
<td>87.3 (52.2)</td>
<td>Highly unfolded conformation is confirmed by NMR, FTIR, SAXS, far-UV CD and gel filtration</td>
</tr>
<tr>
<td>γ-Synuclein (127)</td>
<td>Parkinson’s disease Diffuse Lewy body disease</td>
<td>100 (56.8)</td>
<td>Highly unfolded conformation is confirmed by NMR, FTIR, SAXS, far-UV CD and gel filtration</td>
</tr>
<tr>
<td>Huntingtin (3144; polyQ tract: 16-37 glutamines in norm; &gt;38 glutamines in pathology)</td>
<td>Huntington's disease</td>
<td>35.5 (30.4)</td>
<td>The far-UV CD spectra of poly(Gln) peptides with repeat lengths of 5, 15, 28 and 44 residues were shown to be nearly identical and were consistent with a high degree of random coil structure</td>
</tr>
<tr>
<td>DRPLA protein (1185; polyQ tract: 7-23 glutamines in norm; 49-75 glutamines in pathology)</td>
<td>Hereditary dentatorubral-pallidolysian atrophy</td>
<td>89.5 (84.2)</td>
<td>Aberrant electrophoretic mobility. Apparent molecular mass estimated by SDS-PAGE is ~1.6-fold higher than the predicted molecular mass</td>
</tr>
<tr>
<td>Androgen receptor (919; polyQ tract: 15-31 glutamines in norm; 40-62 glutamines in pathology)</td>
<td>Kennedy's disease or X-linked spinal and bulbar muscular atrophy</td>
<td>53.9 (46.7)</td>
<td>Far-UV CD, gel-filtration, limited proteolysis, ANS binding and urea-induced unfolding studies revealed that the AF1 transactivation domain is in the molten globule state</td>
</tr>
<tr>
<td>Ataxin-1 (816; polyQ tract: 6-39 glutamines in norm; 41-81 glutamines in pathology)</td>
<td>Spinalocerebellar ataxia 1 Neuronal intranuclear inclusion disease</td>
<td>76.8 (73.4)</td>
<td></td>
</tr>
<tr>
<td>Ataxin-2 (1312; polyQ tract: 22-31 glutamines in norm; &gt;32 glutamines in pathology)</td>
<td>Spinalocerebellar ataxia 2</td>
<td>93.8 (76.9)</td>
<td>Ataxin-2 contains two globular domains, Lsm and LsmAD, in an acidic region (amino acid 254-475). The rest of ataxin-2 outside of the Lsm and LsmAD domains is predicted to be intrinsically disordered</td>
</tr>
<tr>
<td>Ataxin-3 (376; polyQ tract: 12-40 glutamines in norm; 55-84 glutamines in pathology)</td>
<td>Spinalocerebellar ataxia 3</td>
<td>52.1 (47.1)</td>
<td>Far-UV CD and NMR spectroscopies suggest that ataxin-3 is only partially folded. The far-UV CD signal of the full-length protein is dominated by the Josephin motif (N-terminal domain 1-198), with the C-terminal portion of the protein making a smaller contribution, consistent with its largely unstructured conformation.</td>
</tr>
<tr>
<td>P/Q-type calcium channel 1A subunit (2505; polyQ tract: 4-16 glutamines in norm; 21-28 glutamines in pathology)</td>
<td>Spinalocerebellar ataxia 6</td>
<td>53.0 (49.3)</td>
<td>Aberrant electrophoretic mobility</td>
</tr>
<tr>
<td>Ataxin-7 (892)</td>
<td>Spinalocerebellar ataxia 7</td>
<td>89.5 (70.2)</td>
<td>Aberrant electrophoretic mobility. Apparent molecular mass estimated</td>
</tr>
</tbody>
</table>
Protein intrinsic disorder in degenerative diseases

Disorder was predicted by two predictors, PONDR® VSL2 and VLXT (given in parenthesis), respectively. PONDR® VSL2 was chosen because of its high accuracy, whereas PONDR® VLXT was chosen because this predictor was shown to be very sensitive for the presence of molecular recognition features, which are disordered polypeptide segments that are predicted to acquire secondary structure upon forming complexes with binding partners.

### 2.4. Relationships between amino acid sequence and fibrillogenesis

Although all proteins can form amyloid fibrils under the appropriate conditions (1), the list of currently known trouble makers (i.e., proteins aggregation of which is directly responsible for the development of various pathologies) is rather short, including 30-40 proteins and protein fragments. Why do some proteins form fibrils under the physiological conditions, whereas others do not? It is clear now that frequently not entire protein is responsible for aggregation and rather aggregation is driven by specific protein regions that contains specific structural or sequence motifs, so-called aggregation-prone regions, APRs, which contribute significantly toward the overall aggregation propensity of the protein (65-72).

It has been pointed out that the propensity of a protein to aggregate can be reliably correlated with a set of simple physico-chemical parameters (73-75). In fact, these specific “aggregation signatures” are concentrated in APRs, which possess a number of recognizable sequence features related to their charge, hydrophobicity, aromaticity, and secondary structure preferences (66, 74-77). All this provides a unique opportunity to identify such aggregation-prone regions by various bioinformatics-based phenomenological approaches that have been specifically developed to predict potential APRs in proteins (72). The existing set of such computational tools is rather large and different research groups are focused on the identification of short (typically 5–9 residues), amyloidogenic regions based solely on the amino acid sequence of a parent protein, whereas other researchers develop tools focused on pattern recognition, 3D profiles, and molecular simulations [reviewed in (72)]. However, one should keep in mind that the presence of APR in a given protein is necessary but not sufficient condition for aggregation since an active APR, which can indeed promote aggregation, should have high intrinsic aggregation propensity, be surface exposed or become exposed upon conformational transition, and facilitate intermolecular interactions (72).

Generally speaking, it is believed that aggregation occurs when protein segments with a high hydrophobicity, a good β-sheet propensity and a low net charge are solvent-exposed so that they can associate, act as nuclei for β-aggregation, and therefore initiate the formation of an intermolecular β-sheet (21, 78-82). In the folded state, such aggregation-prone segments are buried, not exposed to the solvent, and therefore protein does not aggregate. On the other hand, aggregation of many globular proteins occurs during refolding or under conditions in which denatured or partially folded states are significantly populated, i.e. at high concentration or as a result of destabilizing conditions or mutations (5, 14). Based on these findings, the algorithm TANGO was developed to predict β-aggregating stretches in proteins, based on a statistical mechanics algorithm that considers the physico-chemical parameters described above and also takes into account competition between different structural conformations: β-turn, α-helix, β-sheet aggregates and the folded state (83). This algorithm accurately predicted the aggregation propensity of ~250 peptides, including those derived from human disease-related proteins, such as prion protein, lysozyme and beta-microglobulin. It was even able to correctly predict pathogenic as well as protective mutations of the Aβ, human lysozyme and transthyretin, and discriminates between β-sheet propensity and aggregation. Therefore, these data clearly confirmed the model of intermolecular β-sheet formation as a widespread

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Protein</th>
<th>Pathology</th>
<th>APR Prediction</th>
<th>Experimental Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>PolyQ tract: 4-35 glutamines (norm; 36-306 glutamines in pathology)</td>
<td>Spinocerebellar ataxia 17</td>
<td>53.9 (52.3)</td>
<td>Aberrant electrophoretic mobility. Apparent molecular mass estimated by SDS-PAGE is 1.3-fold higher than that calculated from amino acid sequence</td>
<td></td>
</tr>
<tr>
<td>TATA-box binding protein (339; polyQ tract: 25-42 glutamines (norm; &gt;42 glutamines in pathology)</td>
<td>Familial British dementia</td>
<td>29.4 (23.5)</td>
<td>Far-UV CD and NMR spectroscopy revealed that Abri is in the random coil-like conformation at slightly acidic pH</td>
<td></td>
</tr>
<tr>
<td>Abri (34)</td>
<td>Familial Danish dementia</td>
<td>29.4 (23.5)</td>
<td>Far-UV CD revealed that ADan showed mostly random coil structure</td>
<td></td>
</tr>
<tr>
<td>Glial fibrillary acidic protein (432)</td>
<td>Alexander’s disease</td>
<td>82.4 (68.5)</td>
<td>Extremely high susceptibility to proteolysis</td>
<td></td>
</tr>
<tr>
<td>Mitochondrial DNA polymerase (1239)</td>
<td>Alpers disease</td>
<td>37.1 (36.7)</td>
<td>Aberrant electrophoretic mobility</td>
<td></td>
</tr>
<tr>
<td>DNA excision repair protein ERCC-6 (1493)</td>
<td>Cockayne syndrome</td>
<td>56.8 (47.8)</td>
<td>Aberrant electrophoretic mobility</td>
<td></td>
</tr>
<tr>
<td>Survival motor neuron protein (294)</td>
<td>Spinal muscular atrophy</td>
<td>69.7 (60.2)</td>
<td>Aberrant electrophoretic mobility</td>
<td></td>
</tr>
</tbody>
</table>
Protein intrinsic disorder in degenerative diseases

Figure 1. Schematic presentation of the possible mechanisms underlying the spread of α-synuclein pathology and α-synuclein aggregation in PD. In the brain structure on the left, conditions of cellular stress cause α-synuclein to misfold within the neuron (neuron 1), or misfolded α-synuclein is taken up from the extracellular space. Internalized misfolded α-synuclein might be degraded by clearance mechanisms such as the ubiquitin proteasome system, lysosomes and autophagy. Under particular conditions of stress and/or clearance failure, misfolded α-synuclein might not be effectively degraded. Thus, the remaining misfolded α-synuclein might recruit soluble α-synuclein in a seeding mechanism, thereby converting it into misfolded protein, initiating aggregation within neuron 1. The remaining misfolded α-synuclein may also undergo intracellular axonal transport, via fast axonal transport or via slow component b axonal transport within the axon of neuron 1. At the terminal of neuron 1, which is located in brain structure 2, transported (misfolded) α-synuclein might be released by exocytosis, or in exosomes. The α-synuclein released by exocytosis or in exosomes can then be taken up by the surrounding neurons such as neuron 2 (cell-to-cell transfer). The same cascade of events including recruitment of endogenous α-synuclein, seeding and aggregation, clearance and then failure of clearance is proposed to lead first to the formation of α-synuclein aggregates (neurons 1 and 2), and in the end, to the death of the host neuron (dying neuron). Misfolded α-synuclein released into the extracellular space from living neurons or dying cells can activate microglia that take up and degrade misfolded α-synuclein. This figure and figure legend are reproduced from (37) with permission.

In line with these observations, an absolute rate equation was derived from both first principles (i.e., from the detailed analysis of the physicochemical properties of amino acids that are essential for ordered aggregation) and analysis of aggregating sequences designed by a computational approach (84). This model based on physicochemical properties and computational design of β-aggregating peptide sequences was shown to be able to predict the aggregation rate over a large set of natural polypeptide sequences. In addition to evaluating the aggregation rate, the proposed model gave the “amyloid spectrum” of any protein by identifying segments underlying mechanism of protein aggregation (83). Importantly, the application of TANGO also showed that the β-aggregation propensity of α-synuclein as well as membrane-associated proteins was fairly similar, suggesting that β-aggregation was not determined by hydrophobicity and β-sheet propensity alone (82). Importantly, it has been established that globular proteins contained almost three times as much aggregation nucleating regions as IDPs and that the formation of highly structured globular proteins comes at the cost of a higher β-aggregation propensity because both structure formation and aggregation follow very similar physico-chemical rules (82).
Protein intrinsic disorder in degenerative diseases

potentially involved in β-aggregation and was even able to predict the parallel or anti-parallel β-sheet organization in fibrils. An important illustration of the strength of this model was it ability to recognize different β-aggregating segments in mammalian and nonmammalian prion proteins, providing insights into the species barrier for the transmission of the prion disease (84).

In an excellent recent review, Agrawal et al. provided description of many useful computational tools that can identify APRs in a target protein (72). Therefore there is no need to go into a detailed discussion and comparison of these predictors, and the available in silico tools to identify APRs in peptides and proteins are listed below. In addition to mentioned above TANGO, these tools are: AGGRESCAN (http://bioinf.uab.es/aggrescan/) (68, 85), PAGE (84), simple algorithm for sliding average (SALSA) (86), Waltz (http://waltz.vub.ac.be/) (87), predictor of amyloid structure aggregation (PASTA) (http://protein.cribi.unipd.it/pasta/) (88, 89), Zyggregator (http://www-vendruscolo.ch.cam.ac.uk/zyggregator.php) (90-92), Pafig (93), AggreSolve platform, AMYLPRED (http://biophysics.biol.uoa.gr/AMYLPRED) (94), 3D profile method (95), PRE-AMYL (96), FoldAmyloid (97), and spatial aggregation propensity (SAP) (70, 71, 98).

Among other recent developments are AMYLPRED2, which is a consensus method for the prediction of ‘aggregation-prone’ peptides in globular proteins (99), AbAmyloid, which is a method for automatic across-germline prediction of antibody amyloidogenesis from sequence (100), a predictor of protein solubility based on the physicochemical properties of amino acid sequences (101), the NetCSSP tool that implements the latest version of the contact-dependent secondary structure propensity (CSSP) algorithm (102), the BETASCAN algorithm that produces likelihood scores for potential β-strands and strand-pairs based on correlations observed in parallel β-sheets (103), and a tool that combines a coarse-grained physicochemical protein model with a highly efficient Monte Carlo sampling technique to identify amyloidogenic sequences (104).

Also, it was recently shown that four conventional machine learning classifiers, such as Support Vector Machine, Neural network, Decision tree, and Random forest can be trained to provide reliable amyloidogenicity prediction (105). However, one should keep in mind that that the choice of the most accurate and best performing APR prediction tool is challenging, since a complete evaluation of all these algorithms and platforms on standardized dataset(s) in a statistically rigorous manner is still missing (72). Based on a recent comprehensive comparison of the reliability of AGGRESCAN, AMYLPRED, and FoldAmyloid predictions revealed that for these tools a significant reduction of sensitivity was associated with a gain in specificity (106).

A very important recent development is the evaluation of the usefulness of the existing aggregation predictors for predicting the amyloid formation in a biological context (i.e., the ability of these algorithms to predict the effects of mutations on the aggregation of specific proteins, for which in vivo experimental data are available) and the related creation of a supervised, interactive web server AmyloBase (www.unifi.it/scibio/bioinfo/AmyloBase.html) for the deposition of kinetic data of oligomer or amyloid fibril formation and related experimental conditions (107).

Finally, a recent study on the sequence shuffling in the yeast prion proteins Ure2 and Sup35p should be briefly discussed (108). [URE3] and [PSI (+)] being the amyloid-based prions (infectious proteins), are self-propagating amyloid forms of Ure2p and Sup35p proteins of Saccharomyces cerevisiae, respectively. The prion forming ability of Ure2p and Sup35p has been largely localized to short N-terminal domains (so-called Q/N-rich domains). Surprisingly, it has been shown that the sequence of the prion domains is not critical to prion formation, since scrambling sequences of the Ure2p and Sup35p prion domains does not abrogate the ability of the proteins to become prions (108). This observation clearly showed that the amino acid composition of a polypeptide, rather than its specific amino acid sequence, determines the capability of a yeast protein to form amyloid fibrils and thus to become infectious (108).

2.5. Blocking fibrillogenesis

Since protein aggregation is associated with the pathogenesis of several neurodegenerative diseases, it is believed that the prevention of this process may represent an effective therapeutic approach for the treatment of AD, PD, HD, prion diseases, and other neurodegenerative maladies. Therefore, the search for small molecules-inhibitors of protein aggregation represents a very promising area of research. Significant progress is achieved in this field. For example, a number of small molecules have been reported to inhibit Aβ fibrillogenesis, including multiple modulators of Aβ fibrillation (109-127), some of which were able to inhibit Aβ-mediated cellular toxicity (109, 110, 115, 116, 119, 128). In a recent study (129), the anti-oligomer antibody A11 (130) was utilized to find small molecules capable of inhibiting the Aβ aggregation and to characterize the mechanism of action of these Aβ aggregation inhibitors in terms of oligomer and fibril formation. The authors identified a number of small molecules capable of inhibiting oligomer formation, which were grouped into three distinct classes: 17 compounds that inhibited oligomerization but not fibrillation, 5 compounds that inhibited fibrillation but not oligomerization, and 13 compounds that inhibited both oligomerization and fibrillation (129). This study clearly showed that (i) selective inhibition of either Aβ oligomerization or fibrillation is possible, which allows the separate targeting of either species; (ii) the search for fibril inhibitors will only identify a subset of potential oligomer inhibitors since oligomer and fibril formation can be inhibited independently. Finally, it has been pointed out that selective inhibition of specific aggregated species is feasible and useful both for unraveling mechanisms underlying protein fibrillation and for therapeutic testing in models of neurodegeneration (129).
Protein intrinsic disorder in degenerative diseases

Neuritic plaques, neurofibrillary tangles, and neurofilamentous inclusions are hallmark lesions of AD that contain filamentous intraneuronal inclusions of tau protein (131). It is believed that tau filament formation indicates the onset of cytoskeletal disorganization that is characteristic of degenerating neurons and may represent a fundamental pathobiological response of neurons to various insults. Therefore, suppression of tau fibrillation represents an attractive target for the drug development, as selective pharmacological inhibitors of the process may have utility in slowing neurodegeneration associated with the aggregation and fibrillation of this protein. Several small molecules were shown to inhibit tau fibrillation. This includes 3-(2-hydroxyethyl)-2-[6-(2-methoxy-2-benzothiazolylidene)methyl]-1-butyl]-5-methoxy-2-benzothiazolium (N744) (132), phenothiazines (133), anthraquinones (134), polyphenols, porphyrins (135). Recently, a library of ~51,000 compounds was analyzed by high throughput screening to find small molecules inhibitors of heparin-induced tau fibrillation (136). This analysis revealed a set of 11 compounds from eight compound classes: sulfonated dyes, phenothiazines, anthraquinones, benzofurans, porphyrins, quinoxalines, pyrromidotriazines, and a depsidone (136).

Similarly, various small molecules, such as rifampicin (137), dopamine and L-DOPA (138), dopamine-related catecholamines (139), flavanoid baicalein (140, 141), and nicotine (142) were shown to inhibit fibrillation of α-synuclein and to disaggregated the preformed fibrils. Recently, the fibrillation of this protein was successfully inhibited by dendrimers of polyamidoamine (PAMAM) (143). Dendrimers are highly-branched polymers of a well-defined spherical geometry and, thus, are monodisperse in solutions. Their tree-like architecture is synthesized by symmetrical branching from a multifunctional core towards the periphery, via a stepwise and repetitive reaction sequence. Each step of the reaction creates an additional “shell” on the dendrimer surface, called a “generation”. Each generation doubles the molecular weight, as well as both the number of branching points and end-groups at the surface (143). The efficiency of PAMAM dendrimers to inhibit α-synuclein fibrillation and to dissociate fibrils was shown to increase both with generation number and PAMAM concentration (143).

Based on the analysis of the literature data, it has been pointed out that various polyphenols could serve as potential therapeutic agents for the treatment of amyloid-associated diseases, as several small polyphenol molecules were shown to remarkably inhibit the formation of fibrillar assemblies of various proteins in vitro and to suppress their associated cytotoxicity (144). The protective effect of polyphenols against amyloid cytotoxicity in cell culture and primary culture systems, as well as their fibrillation inhibitory activity are frequently attributed to the antioxidative features of polyphenols (145). However, based on structural similarities between various highly efficient polyphenol inhibitors (all of them are composed of at least two phenolic rings with two to six atom linkers, and a minimum number of three OH groups on the aromatic rings), and well-known amyloidogenic dye Congo red, an additional inhibitory mechanism of polyphenol was proposed (144). The mentioned above structural similarities suggest that highly efficient polyphenol inhibitors possess specific three-dimensional conformations that are essential for the non-covalent interaction with β-sheet structures of amyloid fibrils and for the stabilization of the inhibition–protein complex. Additionally, it was proposed that specific interactions between the phenolic compounds in the inhibitor molecules and aromatic residues in the amyloidogenic protein may direct the inhibitor to the amyloidogenic core and facilitate interaction, but interfere with fibril assembly (144).

Significant progress has been achieved in the search for inhibitors of protein aggregation since the time of publication of the original version of this review (146). For example, a comprehensive review by Kurz and Perneeczky provides a systematic view of pharmacological interventions for treatments for AD (147). It is stated that the treatment targets “can be grouped into two major categories. The first category consists of antecedents of Aβ peptide and tau protein deposition including Aβ production, degradation and clearance, tau hyperphosphorylation and aggregation. The second consists of protectors against neuronal dysfunction and premature death such as mitochondrial functioning, nerve growth and regeneration, and neuronal membrane integrity” (147). In another review, Re et al. provide an overview of all small molecules that have been found to interact with Aβ aggregation (148). Among anti-HD therapeutic strategies tested in YAC128 transgenic mice (that express the full-length human HD gene with 128 CAG repeats and serve a unique model for the study of HD) are potentiation of the protective roles of wild-type huntingtin and mutant huntingtin aggregation, transglutaminase inhibition, inhibition of glutamate- and dopamine-induced toxicity, apoptosis inhibition, use of essential fatty acids, and the novel approach of intrabody gene therapy (149). A very promising development in the field of drugs against neurodegenerative diseases is design of multi-targeted agents, or multi-targeted designed drugs (MTDDs) (150). Among such MTDDs are synthetic polycyclic cage compounds, compounds containing the thiazolidinedione moiety, the stilbene scaffold-based compounds, etc. (150).

Some recent advances in the field of inhibitors/modulators of protein aggregation are briefly outlined below. Using a phage display library it was shown that human apolipoprotein A-I (apoA-I) forms non-covalent complexes with Aβ1-42 and affects the morphology of amyloid aggregates formed by Aβ1-42 (151). Furthermore, cerebrospinal fluid from AD patients was shown to contain Aβ1-42/apolipoprotein A-I complexes and apoA-I was able to protect hippocampal neuronal cultures from Aβ-induced oxidative stress and neurodegeneration (151). Galantamine (Reminyl), a currently used drug in the treatment of patients with mild-to-moderate AD was shown recently not only regulate cholinergic transmission, but also efficiently inhibit the Aβ aggregation (152) and protect against oxidative stress induced by Aβ peptide in cortical neurons (153). A prototype nanoparticle-chelator conjugate (Nano-N2PY) was demonstrated to protect human cortical neurons...
Protein intrinsic disorder in degenerative diseases

from Aβ̃-associated oxidative toxicity (154). Various natural and endogenous antioxidants such as polyphenols, coenzyme Q10, and vitamins A, C, and E have shown protective effects against oxidative-induced neuronal death in in vitro and in vivo studies of Parkinson models (155).

An inhibitor of the Rho-associated kinase ROCK, Y-27632, was shown to reduce the huntingtin aggregation in cultured cells and huntingtin-induced neurodegeneration (156). Same Rho-associated kinase ROCK inhibitor Y-27632 was shown to prevent tau hyperphosphorylation in the ischemic rats (157). An increased modification of tau by O-linked N-acetylglucosamine hindered formation of tau aggregates and decreased neuronal cell loss in hemizygous JNPL3 tau transgenic mice (158). Modulation of α-synuclein phosphorylation was proposed as one of the viable strategies for therapeutic interventions in synucleinopathies (159). Inhibition of the ERK phosphorylation by bioavailable bis(thiosemicarbazono)CuII complexes (CuII(bts)) such as diacetyl[bis(μ-thiosemicarbazono)CuII (CuII(atsm))] and glyoxalbis(μ-thiosemicarbazono)CuII (CuII (gtsm)) blocked formation of TDP-43-and HuR-positive RNA stress granules in SH-SY5Y cell ALS model (160).

Tauopathy in the lamprey aquaria model was successfully inhibited by a benzothiazole derivative (E)-2-[[4-(dimethylamino)phenyl][azo]-6-methoxybenzo-thiazole) (161). A series of compounds with a benzophenone scaffold was shown to be effective inhibitors of both the acetylcholinesterase (AChE) and the AChE-induced Aβ aggregation (162). Curcuminoids (polyphenol compounds from turmeric Curcuma longa) may serve as potentially effective therapeutic means to treat various neurodegenerative and protein deposition diseases ranging from AD to Down’s syndrome, to PD, to glaucoma, and to age-related macular degeneration (163). Aggregation of transhyretin and associated cytotoxicity were successfully inhibited by a series of stilbene and dihydrostilbene analogues developed using a substructure combination strategy to generate potent and selective TTR kinetic stabilizers that rescue cells from the cytotoxic effects of TTR amyloidogenesis (164). Methylene blue is able to modulate aggregation of several proteins related to various amyloidogenic diseases including the HD-related protein huntingtin (165). A novel Aβ aggregation inhibitor, SEN1269, was shown to directly bind to monomeric Aβ1-42, produce a concentration-related blockade of Aβ1-42 aggregation, and protect neuronal cell lines exposed to Aβ1-42 (166).

Analysis of the effect of pioglitazone on iron-induced oxidative injury in rat brain revealed that this peroxisome proliferative activated receptor-γ agonist can inhibit iron-induced α-synuclein aggregation and prevent iron-induced apoptosis via both ER and mitochondrial pathways (167). In models of the Machado-Joseph disease, which is the most frequently found dominantly-inherited spinocerebellar ataxia type 3 (SCA3) caused by the proteolysis of the ataxin 3 protein with extended polyglutamine tract, inhibition of calpain activity reduced the size, number and nuclear toxicity of mutant ataxin 3 inclusions, neuronal dysfunction and neurodegeneration (168). Also, application of the histone deacetylase inhibitors (such as valproic acid) in cell and animal SCA3 models suppressed neurodegeneration via alleviating apoptosis and rescuing the hypoacetylation levels of histone H3 and H4 (169).

A rationally designed analogue of and endogenous opioid peptide endomorphin containing efficient β-breaker, α-aminoisobutyric acid, was shown to efficiently interact with Aβ̃ and markedly inhibit the formation of toxic Aβ oligomer and fibril growth (170). Peptides containing the KLVFF sequence, which corresponds to the central region (residues 16-20) of Aβ̃, have been found to be potent inhibitors of Aβ̃ aggregation. The efficiency of inhibition of the Aβ̃ aggregation and the β-amyloid toxicity was significantly increased due to the retro-inversion of these sequences (171). Selective inhibition of transglutaminase was shown to have several very positive effects starting from the normalization of the expression of mitochondrial genes and ~40% of genes that are dysregulated in HD striatal neurons, including chaperone and histone genes, reduction of neuronal degeneration in a Drosophila model of HD, and protection of the mouse HD striatal neurons from excitotoxicity suggesting that transglutaminase inhibition represents a new strategy for treating neurodegeneration (172). The UL97 kinase of the human cytomegalovirus was shown to possess the universal antiaggregation activity, being able to modulate aggregation of several unrelated proteins including polyQ aggregation in vitro and in cellular models of HD and SCA3. This suggested that UL97 might target a key cellular factor that regulates cellular aggregation mechanisms and therefore represents a new means to modulate polyQ aggregation (173). Comparison of the 10 members of the human family of small heat-shock proteins (or HSPB family) for their ability to prevent aggregation of disease-associated proteins with an expanded polyQ stretch revealed that the most active member within the HSPB family was HSPB7, being able to suppress polyQ aggregation and even to prevent polyQ-induced toxicity in cells (174). The reduction of the Insulin/IGF signaling possesses a prominent counter proteotoxic effect and protect models organisms from neurodegeneration-linked toxic protein aggregation (175). Passive vaccination against tau oligomers was proposed as a potential immunotherapeutic approach for the AD treatment (176).

In a direct relation to the topic of this article, a recent review considered various approaches for the design and development of drugs targeting intrinsically disordered proteins, IDPs (177). It was emphasized that the rational development of drugs for inhibition of protein aggregation is complicated by the existence of multiple parallel assembly pathways and the simultaneous existence of various metastable structures in this process (178). Several general strategies in this field are focused on finding small molecules that are able to (177): (i) directly bind to IDRs and block their aggregation by keeping them in the interaction-incompetent conformation; or (ii) interact with IDP/IDR and promote formation and stabilization of the non-toxic and non-amyloidogenic oligomeric species; or
Protein intrinsic disorder in degenerative diseases

(iii) interact with amyloidogenic protein and dramatically accelerate its aggregation to minimize the duration of the toxic oligomer formation stage. An interesting novel idea to find small molecules targeting IDPs and IDR is based on the exploiting the functional misfolding of these proteins (177). Functional misfolding is a mechanism by which IDPs/IDRs are prevented from unwanted interactions with non-native partners via spontaneous formation of a non-interactive cage sequestering interaction-prone preformed fragments (179). Various members of the functionally misfolded ensemble can be stabilized by some small molecules and therefore prevented from establishing native biological interactions. This approach is based on a small molecule binding to a highly dynamic surface created via the transient interaction of preformed interaction-prone fragments (177). In essence, this approach can be considered as an extension of the well-established structure-based rational drug design elaborated for ordered proteins, since if the structure of a member(s) of the functionally misfolded ensemble can be guessed, then it can be used to find small molecules that are potentially able to interact with this structure, utilizing tools originally developed for the rational structure-based drug design for ordered proteins (177).

2.6. Intrinsically disordered proteins, IDPs

2.6.1. Concept

Evidence is rapidly accumulating that many protein regions and even entire proteins lack stable tertiary and/or secondary structure in solution, existing instead as dynamic ensembles of interconverting structures. These naturally flexible proteins are known by different names, including intrinsically disordered (180), natively denatured (181), natively unfolded (182), intrinsically unstructured (183), and natively disordered proteins (184). These proteins are called “intrinsically disordered” from now on. By “intrinsical disorder” it is meant that the protein exists as a structural ensemble, either at the secondary or at the tertiary level. In other words, in contrast to ordered proteins whose 3-D structure is relatively stable and Ramachandran angles vary slightly around their equilibrium positions with occasional cooperative conformational switches, IDPs or intrinsically disordered regions (IDRs) exist as dynamic ensembles in which the atom positions and backbone Ramachandran angles vary significantly over time with no specific equilibrium values and typically undergo non-cooperative conformational changes. To some extent conformational behavior and structural features of IDPs and IDRs resemble those of non-native states of “normal” globular proteins, which may exist in at least four different conformations: ordered, molten globule, pre-molten globule, and coil-like (185-188). Using this analogy, IDPs and IDRs might contain collapsed-disorder (i.e., where intrinsic disorder is present in a molten globular form) and extended-disorder (i.e., regions where intrinsic disorder is present in a form of random coil or pre-molten globule) under physiological conditions in vitro (184, 186, 189). Recently, based on the remarkable sequence and structural heterogeneity of IDPs is has been suggested that structures of IDPs should be described in terms of a continuous spectrum of differently disordered conformations extending from fully ordered to completely structure-less proteins, with everything in between, rather than as a set of several discrete conformations (190).

Recent progress in the field of protein intrinsic disorder is truly remarkable. A simple search of PubMed for “intrinsically disordered protein” gives ~1,270 hits (as of May 08, 2013), with ~950 papers being published since the publication of the original version of this review. Furthermore, during the same time span (from June 01, 2009 to May 08, 2013), various aspects of these proteins were systemized in 150 reviews. Some of these recent reviews are dedicated to specific proteins, such as myelin basic protein (191), measles virus nucleoprotein and phosphoprotein (192), HIV-1 Vif protein (193), p53 (194), α-synuclein (195-198), the genome-linked protein VPG of plant viruses (199), CP12 (200), HMGA (201), EWS-FLI1 fusion protein (202), Hox transcription factor (203), and CBP/p300(204), to name a few. Other reviews talk about roles of intrinsic disorder in various classes and families of IDPs or protein complexes (e.g., colicins (205), cell cycle regulators p21 and p27 (206, 207), neurofilaments (208), LEA proteins (209), steroid hormone receptors (210), protein Ser/Thr phosphatase-1-interacting proteins (211), histone proteins (212), HSF transcription factor family (213), GRASS proteins (214), the nuclear pore complex (215, 216), extracellular matrix (217), 14-3-3 interaction network (218), p53 family (219), etc.). Still other reviews deal with abundance and roles of intrinsic disorder in different pathways and organisms (e.g. HIV-1 proteins (220), viral proteins in general (221), archael proteins (222), plant proteins (223), transcription factors in induced pluripotent stem cells (224), chaperones in neurodegenerative diseases (225), chaperones in general (226), regulators of cell proliferation (227), etc.). Finally, a large cohort of recent reviews covers various general aspects related to IDPs, their functions, and involvement in diseases (146, 177, 190, 228-265).

2.6.2. Experimental techniques for IDP detection

The disorder in IDPs has been detected by several physicochemical methods elaborated to characterize protein self-organization. The list includes but is not limited to X-ray crystallography (266), NMR spectroscopy (184, 267-271), near-UV circular dichroism (CD) (272), far-UV CD (273-276), optical rotatory dispersion (ORD) (273, 276), FTIR (276), Raman spectroscopy and Raman optical activity (277), different fluorescence techniques (278, 279), numerous hydrodynamic techniques (including gel-filtration, viscometry, small angle X-ray scattering (SAXS), small angle neutron scattering (SANS), sedimentation, and dynamic and static light scattering) (278, 279), rate of proteolytic degradation (280-284), aberrant mobility in SDS-gel electrophoresis (285, 286), low conformational stability (278, 287-290), H/D exchange (279), immunochemical methods (291, 292), interaction with molecular chaperones (278), electron microscopy or atomic force microscopy (278, 279), the charge state analysis of electrospray ionization mass-spectrometry (293). For more detailed reviews on methods used to detect intrinsic disorder see (184, 268, 279, 294). The aberrant mobility on SDS-gel electrophoresis will be frequently used in this review as an evidence for the IDP nature of a given protein,
as the anomalous electrophoretic mobility was shown to be one of the characteristic features of IDPs (279, 285, 286). In fact, the apparent molecular masses of IDPs determined by this technique are often 1.2-1.8 times higher than real one calculated from sequence data or measured by mass spectrometry (279, 285, 286). Our analysis revealed that the abnormality degree of the electrophoretic mobility of an IDP is directly proportional to the amount of intrinsic disorder present in its sequence (Uversky, personal communication). It has been suggested that IDPs bind less SDS than “normal” proteins. This explains their abnormal mobility in SDS polyacrylamide gel electrophoresis experiments, resulting in the observed increase in the apparent molecular masses.

It was emphasized that “due to their highly heterogeneous nature and conformational dynamics happening at multiple time-scales, the full spectrum of structural and dynamic characteristics of IDPs cannot be gained by a single tool and clearly requires a multiparametric approach” (295). The current list of biophysical techniques routinely used for the analysis of IDPs/IDRs includes more than 60 approaches. These techniques and the peculiarities of their application for IDP analysis are described in several recent books (296-298). As emphasized in several recent reviews dedicated to this subject (295, 299), the multitude of experimental approaches is absolutely crucial for providing accurate description of structural properties of these highly dynamic and structurally heterogeneous proteins. In fact, the determination of a unique high-resolution structure is not possible for an isolated IDP, multiple complex methods have to be used to obtain experimental constraints on the ensemble of states that is sampled by the intrinsically disordered polypeptide chain. Therefore, IDP-related structural studies typically rely on a host of biophysical methods that can provide information on the overall compactness of IDPs, their conformational stability, shape, residual secondary structure, transient long-range contacts, regions of restricted or enhanced mobility, etc. (295).

High demand for the accurate representation of structural and dynamic features of IDPs serves as a strong motivation for the improvement of existing approaches and the development of new techniques. For example, significant progress has been achieved in the quantitative ensemble descriptions of IDPs from nuclear magnetic resonance spectroscopy, which, due to its specific averaging properties of diverse conformationally dependent observables (269), provides a unique opportunity to map the conformational energy landscape sampled by the protein at atomic resolution (300). As a result, recent years evidenced significant advances in development of calibrated NMR-based approaches to the statistical representation of the conformational behavior of IDPs (300-303) of increasing size and complexity (304). Overall, rational combination of various techniques, such as X-ray crystallography, NMR spectroscopy, and ensemble-modeling strategies employing various experimental measurements, has enabled detailed structural and dynamic characterizations of IDPs and IDPRs (305). The combined use of NMR and small-angle X-ray scattering (SAXS) can provide a detailed structural and dynamic models of IDPs in solution due to the synergy between these two methods (306-308). The advances in the development of complex experimental approaches for structural characterization of IDPs has been accompanied by the development of powerful computational tools to translate experimental results in explicit ensemble representations of IDPs (309). It has been also noted that a four-way synergy between bioinformatics, biophysical experiments, computer simulations, and polymer physics theories represents a crucial foundation for recent advances in quantifying sequence-ensemble relationships for IDPs and IDRs (310).

Among other recent developments in this field is a great progress in single-molecule methods which are new and very powerful tools for dissecting the protein structure and dynamics without associated ensemble averaging (311). Also, electron paramagnetic resonance (EPR) spectroscopy combined with site-directed spin labeling (SDSL) is believed to be amongst the most suitable methods to unravel structure and dynamics of IDPs and IDRs (312).

It has been recently emphasized that various mass spectrometry methods can be used to study IDPs since these techniques “provide a possibility to test biophysical assertions made about why they differ from structured proteins” (313). For example, the uniqueness of the ion mobility-mass spectrometry (IM-MS) is in its ability to examine both absolute conformation(s), populations of conformation, and conformational changes of IDPs (314). Another illustrative example here is the amide hydrogen/deuterium exchange detected by mass spectrometry which has a very broad range of applications and can be used for detecting IDRs in proteins, monitoring coupled folding and binding, and even characterizing protein aggregates and oligomers (315), a direction of special interest in relation to the topic of this review.

2.6.3. Sequence peculiarities of IDPs and predictors of intrinsic disorder

IDPs and IDRs differ from structured globular proteins and domains with regard to many attributes, including amino acid composition, sequence complexity, hydrophobicity, charge, flexibility, and type and rate of amino acid substitutions over evolutionary time. For example, IDPs are significantly depleted in a number of so-called order-promoting residues, including bulky hydrophobic (I, L, and V) and aromatic amino acids (W, F, Y), which would normally form the hydrophobic core of a folded globular protein, and also possess low content of C and N residues. On the other hands, IDPs were shown to be substantially enriched in so called disorder-promoting amino acids: A, R, G, Q, S, P, E, and K (180, 316-318). Many of the mentioned differences were utilized to develop numerous disorder predictors, including PONDR® (316, 319), CH-plot (276), NORSp (320), GlobPlot (321, 322), FoldIndex® (323), IUPred (324), DisoPred (325-327) to name a few. It is important to remember that comparing and combining the results of several predictors on an individual protein of interest or on a protein dataset can
Protein intrinsic disorder in degenerative diseases

provide additional insight regarding the predicted disorder if any exists (328-335).

2.6.4. Natural abundance of IDPs and their biological functions

Application of various disorder predictors to different proteomes revealed that intrinsic disorder is highly abundant in nature and the overall amount of disorder in proteins increases from bacteria to archaea to eukaryota, with over a half of the eukaryotic proteins containing long predicted IDRs (327, 328, 336). These findings were recently consolidated via a comprehensive bioinformatics analysis of completed proteomes of 3484 species from three domains of life (archaea, bacteria and eukaryotes) and from viruses (337). This analysis showed that viruses are characterized by the widest spread of the proteome disorder content (the percentage of disordered residues ranges from 7.3% in human coronavirus NL63 to 77.3% in Avian carcinoma virus). For several organisms, a clear correlation is seen between their disorder contents and habitats. In multicellular eukaryotes, there is a weak correlation between the complexity of an organism (evaluated as a number of different cell types) and its overall disorder content. For both the prokaryotes and eukaryotes, the disorder content is generally independent of the proteome size. However, disorder shows a sharp increase associated with the transition from prokaryotic to eukaryotic cells (337). This suggests that the increased disorder content in eukaryotic proteomes might be used by nature to deal with the increased cell complexity due to the appearance of the various cellular compartments (337). One explanation for this trend is a change in the cellular requirements for certain protein functions, particularly cellular signaling. In support of this hypothesis, an analysis of a eukaryotic signal protein database indicated that the majority of known signal transduction proteins were predicted to contain significant regions of disorder (338).

Although IDPs fail to form unique 3D-structures under physiological conditions, they might carry out important biological functions, the fact which was recently confirmed by several comprehensive studies (180, 183, 184, 186, 189, 270, 276, 286, 294, 338-349). Furthermore, sites of posttranslational modifications (acetylation, hydroxylation, ubiquitination, methylation, phosphorylation, etc.) and proteolytic attack are frequently associated with regions of intrinsic disorder (348). The functional diversity provided by IDRs was suggested to complement functions of ordered protein regions (346-348). IDRs are often encoded by the mRNA regions that are subject for alternative splicing (350). It was postulated that this linkage between alternative splicing and signaling by disordered regions provides a novel and plausible mechanism that could underlie and support cell differentiation, which ultimately gave rise to multicellular organisms in nature (350). The complex “anatomy” of many IDPs/IDRs that often contain multiple, relatively short functional elements contributes to their unique “physiology”, where they are able to be involved in interaction with, regulation of and be controlled by multiple structurally unrelated partners. Given the existence of multiple functions in a single IDP/IDR, and given that each functional element is typically relatively short, alternative splicing could readily generate a set of protein isoforms with a highly diverse set of regulatory elements (350).

IDPs have specific functions that can be grouped into four broad classes: (i) molecular recognition; (ii) molecular assembly; (iii) protein modification; and (iv) entropic chain activities (338). Despite (or may be due to) their high flexibility, IDPs are involved in regulation, signaling and control pathways in which interactions with multiple partners and high-specificity/low-affinity interactions are often requisite (340, 349). In a living organism, proteins participate in complex interactions, which represent the mechanistic foundation of the organism’s physiology and function. Regulation, recognition and cell signaling involve the coordinated actions of many players. To achieve this coordination, each participant must have a valid identification (“ID”) that is easily recognized by the others. For proteins, these “IDs” are often within IDRs (340, 349).

Another very important feature of many IDPs/IDRs is their unique capability to fold under the variety of conditions (180, 183, 185, 267, 270, 276, 286, 294, 338, 340, 345, 349, 351). In fact, the folding of these proteins and regions can be brought about by interaction with other proteins, nucleic acids, membranes or small molecules. It also can be driven by changes in the protein environment. The resulting conformations could be either relatively non-compact (i.e., remain substantially disordered) or be tightly folded. IDPs can form highly stable complexes, or be involved in signaling interactions where they undergo constant “bound-unbound” transitions, thus acting as dynamic and sensitive “on-off” switches (265).

Importantly, spatiotemporal, structural and sequence heterogeneity of IDPs/IDRs also defines their ability to be involved in the formation of the multitude of complexes with a wide range of binding partners. Although binding of IDPs to their partners is often accompanied by the disorder-to-order transitions, many IDPs and IDRs preserve significant amount of disorder even in their bound states (265). In fact, these IDPs form so-called disordered, dynamic, or fuzzy complexes with ordered proteins (352-357), other disordered proteins (358-360), or biological membranes (361, 362). Overall, IDPs/IDRs can form static, semi-static and dynamic complexes (265). Static and semi-static binding modes range from the interaction-induced gaining of local structure on the surface of a binding partner to folding of a whole molecule, and from wrapping around the binding partner to penetrating deep inside the binding partner. IDPs can participate in one-to-many and many-to-one interactions, where one IDR binds to multiple partners potentially gaining very different structures in the bound state, or where multiple unrelated IDPs/IDRs bind to one partner (265). Binding functions of IDPs and IDRs are controlled by various means, such as numerous posttranslational modifications and alternative splicing (265).
Protein intrinsic disorder in degenerative diseases

3. FIRST D\textsuperscript{3}: ABUNDANCE OF IDPs IN NEURODEGENERATIVE DISEASES AS EVIDENT FROM THE BIOINFORMATICS ANALYSES

Because of the fact that IDPs play a number of crucial roles in numerous biological processes, it was not too surprising to find that some of them are involved in human diseases. An incomplete list of human neurodegenerative diseases associated with IDPs includes AD (deposition of amyloid-\(\beta\), tau-protein, \(\alpha\)-synuclein fragment NAC (363, 364); Niemann-Pick disease type C, subacute sclerosing panencephalitis, argyrophilic grain disease, myotonic dystrophy, and motor neuron disease with neurofibrillary tangles (NFTs) (accumulation of tau-protein in form of NFTs (365)); Down’s syndrome (nonfilamentous amyloid-\(\beta\) deposits (366)); PD, dementia with Lewy body (LB), diffuse LB disease, LB variant of AD, multiple system atrophy (MSA) and Hallervorden-Spatz disease (deposition of \(\alpha\)-synuclein in a form of LB, or Lewy neurites (LN) (367)); prion diseases (deposition of PrPSC (368)); and a family of polyQ diseases, a group of neurodegenerative disorders caused by expansion of GAC trinucleotide repeats coding for polyQ in the gene products (369).

Table 1 and Figure 2 illustrates that some individual proteins involved in human neurodegenerative diseases are either completely disordered or contain long disordered regions. Figure 2 represents the results of the comparison of the compositions of proteins from Table 1 with the composition of ordered proteins from PDB. The corresponding data for the DisProt (370) are shown for comparison. Calculations were done using a normalization procedure elaborated for analysis of IDPs (180, 371). In brief, compositional profiling is based on the evaluation of the \((C_{\text{S1}} - C_{\text{S2}})/C_{\text{S2}}\) values, where \(C_{\text{S1}}\) is a content of a given residue in a set of interest (proteins associated with neurodegenerative diseases or typical IDPs from DisProt), whereas \(C_{\text{S2}}\) is the corresponding value for the set of ordered proteins. In this presentation, negative values correspond to residues which are depleted in a given dataset in comparison with a set of ordered proteins, whereas the positive values correspond to the residues which are over-represented in the set.

Figure 2 shows that in general all proteins in Table 1 are highly different from typical ordered proteins and generally follow the trend for IDPs (with some exceptions). Proteins associated with neurodegenerative diseases are in general depleted in major order-promoting residues. This includes C, W, I, Y, F, V and N. They are highly enriched in the major disorder-promoting residues (Q, S, R, and P). There are also some deviations from the behavior of “typical” disordered proteins. This includes the
Protein intrinsic disorder in degenerative diseases

Figure 3. Abundance of intrinsic disorder in proteins associated with neurodegenerative diseases. Percentages of disease-associated proteins with \( \geq 30 \) to \( \geq 100 \) consecutive residues predicted to be disordered. The error bars represent 95% confidence intervals and were calculated using 1,000 bootstrap re-sampling. Corresponding data for signaling and ordered proteins are shown for the comparison. Analyzed protein sets included 1,786 proteins associated with cancer, 689 proteins involved in the neurodegenerative diseases, 2,329 proteins involved in cellular signaling, 1,138 non-homologous ordered proteins from PDB Select 25 (this dataset contained only the ordered parts of the proteins), and 53,630 non-redundant eukaryotic proteins from SWISS-PROT.

high abundance of L and H and the depletion in T, D, and K. This suggests that proteins listed in Table 1 are in general characterized by a high level of intrinsic disorder.

This fact raises the question of how abundant are the IDPs in various neurodegenerative conditions. To answer this question, a set of 689 proteins related to neurodegenerative diseases was collected and analyzed using an approach elaborated to analyze the abundance of intrinsic disorder in cancer-related proteins (351). In that study, 79% of cancer-associated and 66% of cell-signaling proteins were found to contain predicted regions of disorder of 30 residues or longer (351). In contrast, only 13% of proteins from a set of proteins with well-defined ordered structures contained such long regions predicted to be disordered by PONDR® VLXT. In agreement with these bioinformatics studies, the presence of intrinsic disorder has been directly observed in many cancer-associated proteins (351).

The overall results of the analogous analysis for proteins associated with neurodegenerative disease are shown in Figure 3, which represents percentages of proteins with \( \geq 30 \) consecutive residues predicted to be disordered by PONDR® VSL2 in various datasets, including cancer-related proteins, signaling proteins, ordered proteins from PDB, eukaryotic proteins and proteins involved in various neurodegenerative diseases. This figure illustrates that intrinsic disorder is highly prevalent in neurodegenerative disease-related proteins, being comparable with that of signaling and cancer-related proteins and significantly exceeds the level of intrinsic disorder in eukaryotic proteins from SWISS-PROT and in non-homologous, structured proteins from the PDB. Thus, intrinsic disorder is very common in neurodegeneration-associated proteins. To further illustrate this concept, Table 1 represents some of the IDPs and their corresponding neuropathological conditions. Many of these proteins were structurally characterized and experimental evidence on the presence of intrinsic disorder in some of these proteins is also listed. Table 1 shows that there is a great agreement between experimental and computational data. Finally, the results of the disorder prediction by two predictors, PONDR® VSL2 and VLXT, are shown. Figure 4 represents plots of the PONDR® VSL2 predicted disorder distribution within the sequences of 20 neurodegenerative disease-related IDPs. It clearly shows that these proteins are very divergent: their length range from 34 to 3144 amino acids, the amount of predicted disorder range from 16.7 to 100%, and the profiles of disorder distribution are very different.
Protein intrinsic disorder in degenerative diseases

Figure 4. Distribution of intrinsic disorder in neurodegeneration-related proteins as predicted by PONDR VSL2. A, Aβ; B, Tau protein; C, Prion protein; D, α-Synuclein; E, β-Synuclein; F, γ-Synuclein; G, Huntingtin; H, DRPLA protein (atrophin-1); I, Androgen receptor; J, Ataxin-1; K, Ataxin-2; L, Ataxin-3; M, P/Q-type calcium channel α1A subunit; N, Ataxin-7; O, TATA-box-binding protein; P, ABri; Q, Glial fibrillary acidic protein; R, Mitochondrial DNA polymerase γ; S, DNA excision repair protein ERCC-6; T, Survival motor neuron protein. Shaded areas in each plot correspond to the scores associated with intrinsic disorder. Details of the analyzed proteins are listed in Table 1.

Therefore, computational analysis showed that the majority of the proteins involved into the pathogenesis of neurodegenerative disease are intrinsically disordered. Subsequent sections consider illustrative examples of some of the most important neurodegenerative IDPs and their corresponding diseases.

4. α-SYNUCLEIN AS AN ILLUSTRATIVE EXAMPLE OF IDPS RELATED TO THE NEURODEGENERATION

Before considering various neurodegeneration-related IDPs and their related maladies, some characteristic structural features of these proteins should be introduced. This is done using an illustrative example of α-synuclein, which is the one of the most thoroughly studied IDPs. A brief description of structural properties of this protein is presented below. It is also pointed out that this protein is likely to maintain its intrinsically disordered structure even being located in a highly crowded environment of a living cell.

4.1. Structural properties of α-synuclein: Protein-chameleon

α-Synuclein, a protein that links various synucleinopathies, is one of the most studied IDPs. It possesses little or no ordered structure under the "physiological" conditions in vitro (i.e., conditions of neutral pH and low to moderate ionic strength) (372). For example, at neutral pH α-synuclein is characterized by far-UV CD and FTIR spectra typical of a substantially unfolded polypeptide chain with a low content of ordered secondary structure (Figures 5A and 5B). This includes: the characteristic minimum in the vicinity of 196 nm and the
absence of bands in the 210-230 nm region in far-UV CD spectrum and broad band at 1650 cm$^{-1}$ in the FTIR spectrum. Deconvolution of the FTIR spectra followed by curve fitting revealed that the majority of the molecule (~70%) is disordered (372). The hydrodynamic properties of this protein are in a good agreement with the results of the far-UV CD and FTIR studies and show that $\alpha$-synuclein, being essentially expanded, does not have a tightly packed globular structure, but is slightly more compact than expected for a random coil (372). This follows from the comparison of values of the measured Stokes radius, $R_g$, with those calculated for a completely unfolded polypeptide chain of the appropriate molecular mass (294, 345, 373, 374). It has been shown that the Stokes radius measured for $\alpha$-synuclein was notably lower than the corresponding calculated value (31.8±0.4 vs. 34.3 Å, (345)). This conclusion was confirmed by measurement of the $\alpha$-synuclein $R_g$ in the presence of 8 M urea, where the protein behaved as a random coil ($R_g$=34.5±0.4 Å, (345)).

SAXS is a very useful method for the investigation of conformation, shape and dimensions of biopolymers in solution. Analysis of the scattering curves using the Guinier approximation provides the radius of gyration, $R_g$. Scattering data in the form of Kratky plots provides information about the globularity (packing density) and conformation of the protein (375): for a native globular protein this plot has a characteristic maximum, whereas unfolded and partially folded polypeptides have significantly different-shaped Kratky plots. Figure 4C represents Kratky plots for a typical globular protein (Staphylococcal nuclease) and $\alpha$-synuclein at variety of conditions. Figure 5C clearly shows that $\alpha$-synuclein lacks a well-developed globular structure at both conditions studied (pH 7.5 and pH 3.0). In fact, the profile of the Kratky plot at neutral pH is typical for a random coil conformation, whereas that at pH 3 shows changes consistent with the development of the beginnings of a tightly packed core.

The radius of gyration of a completely unfolded polypeptide, $R_g^{U}$, may be estimated from the corresponding Stokes radius, $R_g$, using the relation $R_g^{U}/R_g=1.51$ (376). The observed $R_g$ value for $\alpha$-synuclein at neutral pH (40±1 Å) is smaller than that estimated for a random coil conformation for a protein of this size (52 Å), indicating that the natively unfolded conformation of this protein is more compact than that of a random coil (186, 372, 377-379). Finally, the profile of the Kratky plot at neutral pH was typical for a random coil conformation (372, 377-379). Thus, at neutral pH $\alpha$-synuclein was shown to be essentially disordered, but slightly more compact than a random coil. Based on the results of pulsed-field gradient NMR (which allows an estimation of the hydrodynamic radii), it has been concluded that $\alpha$-synuclein is slightly collapsed (380). In agreement with this conclusion, a high resolution NMR analysis of the protein revealed that $\alpha$-synuclein is largely unfolded in a solution, but exhibits a region between residues 6 and 37 with a preference for helical conformation (381). Interestingly, Raman optical activity spectra indicate that $\alpha$-synuclein contains some helical poly-(L-proline) II-like conformation (382).

$\alpha$-Synuclein, with its high propensity to aggregate, represents an ideal model for the amyloidogenic IDP and the molecular mechanisms underlying the amyloidogenesis of this protein were intensively studied. It has been shown that $\alpha$-synuclein partially folds at acidic pH and high temperature; i.e., under conditions that enhanced dramatically the propensity of the protein to form amyloid-like fibrils (372). These behaviors are illustrated by Figure 5 which shows that $\alpha$-synuclein adopts a partially folded conformation at acidic pH or at high temperatures (cf. (186, 372, 377-379)). At neutral pH the protein possesses a far-UV CD spectrum typical of an unfolded polypeptide chain (Figure 5A). The spectrum has an intense minimum in the vicinity of 196 nm, with the absence of characteristic bands in the 210-230 nm region. However, as the pH is decreased (or temperature increased) changes were observed in the shape of the spectrum. Figure 5A shows that the minimum at 196 nm becomes less intense, whereas the negative intensity of the spectrum around 222 nm increases, reflecting pH-induced formation of secondary structure. Figure 4B compares the FTIR spectra of $\alpha$-synuclein measured at pH 7.5 and pH 3.0. The FTIR spectrum of $\alpha$-synuclein at pH 7.5 is typical of a substantially unfolded polypeptide chain, whereas a decrease in pH leads to significant spectral changes, indicative of increased ordered structure. The most evident change is the appearance of a new band in the vicinity of 1626 cm$^{-1}$, which corresponds to $\beta$-sheet. This means that at acidic pH natively unfolded $\alpha$-synuclein is transformed into a partially folded conformation with a significant amount of $\beta$-structure (372, 377-379). Furthermore, Figure 5D shows that a decrease in pH leads to a large blue shift of the 8-anilino-1-naphthalenesulfonic acid (ANS) fluorescence maximum (from ~515 to ~475 nm, open triangles in Figure 5D), reflecting the pH-induced transformation of the natively unfolded $\alpha$-synuclein to the partially folded compact conformation. Figure 4D shows that the pH-induced structural transitions observed by ANS fluorescence and CD change simultaneously in a rather cooperative manner. This means that protonation of $\alpha$-synuclein results in transformation of the natively unfolded protein into a conformation with a significant amount of ordered secondary structure and with affinity for ANS. The position of the transition (between pH 5.5 and 3.0) indicates that protonation of one or more carboxylates is responsible for the structural change. Finally, pH-induced transition from unfolded to partially folded conformation was shown to be completely reversible (Figure 5D, open and solid symbols, (372, 377-379)). Hydrodynamic methods revealed that pH-induced formation of partially folded conformation is accompanied by substantial decrease in hydrodynamic dimensions ($R_g=27.9±0.4$ and $R_g=30±1$ Å). Furthermore, changes in the profile of the Kratky plot at pH 3 were consistent with the development of the beginnings of a tightly packed core (See Figure 5C) (372, 377-379).

Figure 5E represents the temperature-dependence of [0]$_{222}$ and shows that increase in temperature induced formation of secondary structure in $\alpha$-synuclein (383). The major spectral changes occurred over the range of 3 to 50°C. Further heating lead to a less pronounced effect.
Interestingly, Figure 5E shows that the structural changes
duced in α-synuclein by heating were completely
reversible (cf. open and filled symbols). These data indicate
that high temperatures induce a reversible transition of α-
synuclein to a partially folded intermediate. This
intermediate has a similar CD spectrum to that induced by
low pH (383).

Conformational behavior of α-synuclein under the
variety of environmental conditions revealed that
structure of this protein is extremely sensitive to the
environment. It adopts a variety of structurally unrelated
conformations including the substantially unfolded state,
an amyloidogenic partially folded conformation, and
different α-helical or β-structural species folded to a
different degree, both monomeric and oligomeric (372).
Furthermore, it might form several morphologically
different types of aggregates, including oligomers
(spheres or doughnuts), amorphous aggregates, and
amyloid-like fibrils (372). Based on this astonishing
conformational behavior the concept of a protein-
chameleon was proposed, according to which the
structure of α-synuclein to a dramatic degree depends
on the environment and the choice between different

Figure 5. Structural properties and conformational behavior of α-synuclein. A. Far-UV CD spectra measured at different pH. B.
FTIR spectra measured for natively unfolded, partially folded and fibrillar forms. C. Kratky plots for native unfolded at pH 7.5
(1) and partially folded α-synuclein at pH 3.0 (2) in comparison with the typical globular protein, staphylococcal nuclease (3). D.
conformations is determined by the peculiarities of protein surroundings (372).

4.2. α-Synuclein maintains disordered structure inside crowded environment

The cell’s interior is crowded with small and large molecules (384, 385), with the concentration of macromolecules, including proteins, nucleic acids, and carbohydrates, being as high as 400 g/L (386), and with considerably restricted amounts of free water (384, 386-391). The volume occupied by the macromolecular co-solutes is unavailable to other molecules, giving rise to the so-called excluded volume effects (387, 392). Although this phenomenon is often neglected, some researchers analyze the effects of macromolecular crowding on protein function, structure and conformational behavior (384, 386-391) using conditions that mimic the “inside the cell” environment; i.e., in the presence of concentrated solutions of model “crowding agents” such as polyethylene glycol, dextran, Ficoll or inert proteins (392, 393). The excluded volume effects induced by such crowded environments were shown to affect the behavior of biological macromolecules (390, 394-396), and protein-protein interactions (380, 390, 391), including modulation of the rate and extent of amyloid formation (393, 397, 398).

Despite expectation that crowding should promote folding of α-synuclein, this protein was shown to remain mostly unfolded under the variety of artificial crowding conditions (399-401). Based on the measurement of NMR spin relaxation parameters of α-synuclein in the presence of up to 400 g/L of Ficoll 70 or Dextran 70 it has been concluded that this protein not only preserved its overall intrinsically disordered structure in crowded environment, but also retained its segmental motions on the nanosecond timescale clearly indicating that this IDP exists and functions as a monomeric protein (404). Later, this conclusion was confirmed by several other research groups (408-410). This is a very important observation, which suggests that some IDPs, including α-synuclein, can maintain their disordered structure even in the highly crowded environment of a living cell.

4.4. α-Synuclein as a disordered hub

α-Synuclein is known to be involved in multiple physical interactions with various proteins. In the case-by-case studies, α-synuclein was shown to interact with at least 50 proteins (367), whereas a recent proteomic analysis using a SILAC technique (stable isotope labeling by amino acids in cell culture) identified 587 proteins involved in the formation of complexes with α-synuclein in the dopaminergic MES cells, with 141 proteins displaying significant changes in their relative abundance (increase or decrease) after the MES cell were treated with rotenone (411).

Recent bioinformatics studies showed that the common structural feature of many hub proteins is their intrinsically disordered nature or their ability to interact with intrinsically disordered partners (459-463). Due to its intrinsically disordered nature and the aforementioned multitude of interactions, α-synuclein most definitively serves as disordered hub protein. This is further illustrated by the α-synuclein-centered interaction network shown in Figure 6. This figure represents the results of the analysis of α-synuclein interactome using the STRING database, which is the online database resource Search Tool for the Retrieval of Interacting Genes providing both experimental and predicted interaction information (464). STRING database acts as a ‘one-stop shop’ for all information on functional links between proteins, and version 9.0 of STRING (accessible at http://string-db.org) covers more than 1,100 completely sequenced organisms, including Homo sapiens. This tool produces the network of predicted associations for a particular group of proteins. The network nodes are proteins, whereas the edges represent the functional associations evaluated based on the experiments, search of databases and text mining. Different line colors of edges represent different types of evidence for the interactions.
Protein intrinsic disorder in degenerative diseases

Figure 6. α-Synuclein is an intrinsically disordered hub. Interactome of α-synuclein is analyzed using the STRING database (accessible at http://string-db.org) generates the network of predicted associations for a particular group of proteins. The network nodes are proteins, whereas the edges represent the functional associations evaluated based on the experiments, search of databases and text mining. Different line colors of edges represent different types of evidence for the association, such as experimental/biochemical data (pink lines), association in curated databases (blue lines), and co-mentioned in PubMed abstracts (dark yellow lines).

4.5. A place of the helical α-synuclein tetramers within a complex structural image of protein-chameleon

Based on the considered above facts and on the wealth of related publications it seems to be clear that α-synuclein should be considered as a classic, well-studied example of IDPs with great clinical potential. In fact, this was a dominating view for about 15 years (since the first publication describing the “natively unfolded” nature of this protein (182)). However, this concept was challenged by two studies published in 2011, in which it was claimed that physiologically, α-synuclein occurs as helical folded tetramer that resists aggregation (465), and that heterologously expressed form of this protein with a 10-residue N-terminal extension (GPLGSPEFPG) forms a stable tetramer in the absence of lipid bilayers or micelles (466). These papers attracted significant attention of the field not only because they were published in high profile journals, but mostly because they stated that everything done during these 15 years of the intensive research by many groups around the globe was wrong. The reason for this discrepancy was stated to be rooted in wrong protocols for α-synuclein purification (465): “α-Synuclein has long been defined as a 'natively unfolded' monomer of about 14 kDa that is believed to acquire α-helical secondary structure only upon binding to lipid vesicles. This concept derives from the widespread use of recombinant bacterial expression protocols for in vitro studies, and of overexpression, sample heating and/or denaturing gels for cell culture and tissue studies. In contrast, we report that endogenous α-synuclein isolated and analyzed under non-denaturing conditions from neuronal and non-neuronal cell lines, brain tissue and living human cells occurs in large part as a folded tetramer of about 58 kDa... Whereas recombinantly expressed monomers readily aggregated into amyloid-like fibrils in vitro, native human tetramers underwent little or no amyloid-like aggregation.” Aside from these assembly and purification issues, there was also one molecular difference between the purified samples of Selkoe and colleagues and previous studies, indicative of modification to the monomer by an acetyl group (467).

Challenged by these bold statements, several groups undertook comprehensive in depth analysis of various α-synuclein samples under the variety of conditions (409, 410, 468-471). These studies have shown that both non-acetylated and acetylated α-synuclein purified under mild conditions predominantly exists as a monomeric intrinsically disordered conformational ensemble (467).
Figure 7. α-Synuclein as a protein-chameleon. Major structural forms of α-synuclein are shown: (0) Intrinsically disordered coil-like structure of α-synuclein in the non-bound form. This structural ensemble was generated by the Monte Carlo conformational search constrained by smFRET measurements (Credits: E. Rhoades, Yale University; Source: http://www.yale.edu/rhoadeslab/research.html); (1) A partially folded α-synuclein when bound to a micelle of the detergent sodium lauroyl sarcosinate (PDB ID: 2KKW); (2) Proposed α-synuclein tetramer structure based on electron microscopy reconstruction and paramagnetic relaxation enhancement; (3) Pentamer of α-synuclein 4ns molecular dynamics conformers on the membrane (Credit: I. Tsigelny, Y. Sharikov, M. Miller, and E. Masliah, SDSC/UCSD; Source: San Diego Supercomputer Center, UC San Diego) (4) α-Synuclein bound to large unilamellar vesicles (modified from (893)). This figure is reproduced with permission from (894).

The logical explanation for these discrepancies lies within the mentioned chameleon nature of α-synuclein. Since structure of this protein (and probably all IDPs in general) is extremely sensitive to its environment, and since, depending on the peculiarities of its environment, α-synuclein can populate structurally different monomers, various soluble oligomeric species with different morphologies and toxicities, and also can form insoluble fibril or amorphous aggregate with diverse morphologies (372), the mentioned soluble, aggregation-resistant, and α-helical tetramer is likely to be just one of many potential structures realized under the umbrella of the dynamic conformational ensemble (467). This idea is further illustrated by Figure 7 showing some of the structures attainable by α-synuclein.

5. NEURODEGENERATION-ASSOCIATED INTRINSICALLY DISORDERED PROTEINS AND CORRESPONDING MALADIES

5.1. Amyloid β-protein and Alzheimer’s disease

AD is the most prevalent age-dependent dementia, causing cognitive decline among people of age 65 and older. It currently affects 4.5 million Americans and is projected to afflict 13.2 million by the year 2050 in the US alone (472). AD ranks third in total health care cost after heart disease and cancer. The national direct and indirect annual cost of AD approaches 100 billion dollars per year (473).

AD was described for the first time in 1907 by a German physician Alois Alzheimer (474). AD is the most common aging-related neurological disorder, which
constitutes about two thirds of cases of dementia overall (475, 476) and is characterized by slow, progressive memory loss and dementia due to a gradual neurodegeneration particularly in the cortex and hippocampus (477). The clinical hallmarks are progressive impairment in memory, judgment, decision making, orientation to physical surroundings, and language (478). From the initial symptoms, disease progression can last up to 25 years, although typically the duration ranges from 8 to 10 years.

Sporadic AD is a disease of the elderly; most patients are diagnosed after 65 years of age. About 10 % of AD cases present under age 65 and have been referred to as having early onset AD. Three causative autosomal dominant mutations have been described – the amyloid β-protein precursor (APP) gene mutation on chromosome 21, the presenilin 1 gene mutation on chromosome 14 and the presenilin 2 gene mutation on chromosome 1. These autosomal dominant forms comprise only about 2% of all AD (479). Having an extra copy of the APP gene, as in Down’s patients (trisomy 21), also leads to early pathological and clinical changes of AD.

AD is characterized biochemically by the accumulation of two types of proteinaceous inclusions, extracellular amyloid deposits, senile plaques, in the cerebral cortex and vasculature and intracellular NFTs (paired helical filaments, PHFs) (480). Amyloid is a descriptive term for proteinaceous deposits that stain with Congo red and thioflavin S and demonstrate birefringence in polarized light. Amyloid deposits in AD contain the amyloid β-protein (Aβ), which is a 40–42 residue peptide, produced by endoproteolytic cleavage of the APP. PHFs are assembled from a hyperphosphorylated form of the microtubular protein tau (see next section).

APP, the parent molecule of Aβ, plays a role in synaptic stabilization and plasticity, regulation of neuronal survival, neuritic outgrowth and cell adhesion (481, 482). Nexin-2, a secreted form of APP, inhibits coagulation factor Xla (483, 484). C-terminal fragment of APP originating after the γ-secretase cleavage mediates nuclear signaling and modulate gene expression (485-488). The Aβ fragment of the APP protein is a byproduct of APP processing. The normally prevailing α-secretase-mediated APP processing splits the large APP molecule in the middle of the Aβ sequence and does not produce pathogenic Aβ species. However, alternative cleavage by the β- and γ-secretases results in generation of the pathogenic Aβ fragment. Depending on the exact site of action of γ-secretase, several Aβ peptides with 39-43 amino acids are produced (489). The longer moieties are more amyloidogenic (130). Although β- and γ-secretase are active throughout the lifespan, plaques rarely form in young individuals, but after the age of 60 nearly all elderly develop some Aβ deposits (490, 491).

Many lines of evidence support the crucial role of Aβ in AD. Aggregated forms of the Aβ peptide with amyloid-like cross-β structure are neurotoxic to cortical cell cultures (492-495). Some of the Aβ derived diffusible ligands (small Aβ aggregates) kill mature neurons at nanomolar concentrations and cause neurological dysfunction in the hippocampus (496). The two major Aβ peptides are the 40-residue Aβ1-40 and the 42-residue Aβ1-42, which differ in the absence or presence of two extra C-terminal residues (Ile41-Ala42). The N-terminal (residues 1-28) residues comprise a hydrophilic domain with a high proportion of charged residues (40%), whereas the C-terminal domain (residues 29-40 or 29-42) is completely hydrophobic and is presumably associated with the cell membrane. Although the Aβ1-40 and Aβ1-42 peptides are ubiquitous in biological fluids of humans (at an approximate ratio of 9:1), it is thought that the longer Aβ1-42 is more pathogenic, due to its higher quantities in the amyloid plaques of sporadic AD cases, its even higher quantities in patients afflicted with early onset AD (497, 498), and because of the greater in vitro tendency of the Aβ1-42 to aggregate and precipitate as amyloid (499, 500). Fibrillation of Aβ is associated with the development of the cascade of neuropathogenic events, ending with the appearance of cognitive and behavioral features typical of AD.

Aβ appears to be unfolded at the beginning of the fibrillation under physiological conditions. NMR studies have shown that monomers of Aβ1-40, or Aβ1-42 possess no α-helical or β-sheet structure (501); i.e., they exist predominantly as random coil-like highly extended chains. Partial refolding to the pre-molten globule-like conformation has been detected at the earliest stages of Aβ fibrillation (501).

Besides AD, Aβ aggregation was implicated in several other neurodegenerative diseases (see Table 1). For example, the E22Q mutation of Aβ is associated with the rare disorder, hereditary cerebral hemorrhage with amyloidosis-Dutch type (HCHWA-D). HCHWA-D is characterized by severe cerebral amyloid angiopathy (CAA), which is characterized by extensive amyloid deposition in the small leptomeningeal arteries and cortical arterioles, leading to hemorrhagic strokes of mid-life onset, dementia and an early death of those afflicted in their fifth or sixth decade. Therefore, this disorder is an autosomal dominant form of vascular amyloidosis restricted to the leptomeninges and cerebral cortex. CAA severity tends to increase with age (502). In HCHWA-D, parenchymal Aβ deposition is enhanced, with non-fibrillar membrane-bound Aβ1-42 deposits evolving into relatively fibrillar diffuse plaques variously associated with reactive astrocytes, activated microglia, and degenerating neurites (502). Although silver stain-positive, “senile plaque-like” structures found in the HCHWA-D brain were immunopositive for Aβ, yet these lesions lacked the dense amyloid cores present in typical AD plaques (503). No NFTs are present in this disorder. The total Aβ production is not affected by E22Q mutation. However, the proteolytic degradation of Aβ and its transport across the blood–brain barrier as well as the Aβ1-42/Aβ1-40 ratio are altered. Aβ E22Q aggregates faster and fibrils formed by this variant are more stable than amyloid-like fibrils produced by the wild-type Aβ (502).
5.2. Tau protein in Alzheimer’s disease and other tauopathies

The tau gene is located on chromosome 17. It encodes for a protein with four 31-32 amino acid tandem repeats close to its C-terminus. Tau protein is a vital structural element of the microtubular transport system in the nervous system. Its aggregation is implicated in AD and several other diseases collectively known as tauopathies (see Table 1).Tau protein represents a family of isoforms migrating as close bands of 55-62 kDa in SDS gel electrophoresis. Heterogeneity is due in part to alternative mRNA splicing. The tau primary transcript contains 16 exons. Exons 1, 4, 5, 7, 9, 11, 12 and 13 are constitutive exons, exon 14 is part of the 3′ untranslated region of tau mRNA, whereas exons 2, 3 and 10 are alternatively spliced (504). The alternative splicing of these three exons produces six combinations (2−3−10−; 2+3−10−; 2+3+10−; 2−3−10+; 2+3−10+; 2+3+10+), and in the human brain, the tau primary transcript gives rise to six mRNAs (505, 506). Thus, in the human brain, the tau proteins constitute a family of six isoforms with the range from 352-441 amino acids. They differ in either no, one or two inserts of 29 amino acids at the N-terminal part (exon 2 and 3), and three or four repeat-regions at the C-terminal part exon 10 missing. The longest isoform in the CNS (441 amino acids total) has four repeats (R1, R2, R3 and R4) and two inserts, while the shortest isoform (352 amino acids total) has three repeats (R1, R3 and R4) and no insert (507, 508). In normal cortex the three and four-repeat forms are equally expressed. In tauopathies the ratio of isoforms is changed. AD is the only dementia with both three- and four-repeat tau (509). Furthermore, it has been shown that pathological tau proteins in different tauopathies are characterized by different electrophoretic patterns, representing the bar code of tauopathies based on variation in distribution of the pathological tau bands at 60, 64, 69 and 74 kDa (510).

In vitro, tau binds to microtubules, promotes microtubule assembly, and affects the dynamic instability of individual microtubules (511-515). In situ, tau is highly enriched in the axons (516). In living cells and brain tissue, tau protein has been estimated as comprising 0.025-0.25% of total protein (517, 518). On the basis of its in vitro activity and its distribution, it is believed that tau regulates the organization of neuronal microtubules. Interest in tau dramatically increased with the discovery of its aggregation in neuronal cells in the progress of AD and various other neurodegenerative disorders, especially frontotemporal dementia (519, 520). In these cases specific tau-containing NFTs or PHFs are formed (520). Hyperphosphorylation was shown to be a common characteristic of pathological tau (521). Hyperphosphorylated tau isolated from patients with AD was shown to be unable to bind to microtubules and promote microtubule assembly. However, both of these activities were restored after enzymatic dephosphorylation of tau protein (522-525). Although tau inclusions can be stained with hematoxylin-eosin and amyloid stains, they are much easier visualized after silver impregnation. The most sensitive and specific method is tau immunohistochemistry. There are three types of tau deposits in AD – NFTs, neurofibril threads, and dystrophic neurites.

NFTs are composed of 22 nm PHF and each PHF is composed of 8-14 tau monomers (526). They commonly affect the pyramidal cortical neurons and assume a flame-like shape. Extracellular NFTs are rare and are referred to as ghost tangles. They are presumed to be the remnants of dead neurons and are most commonly seen in the hippocampus. When surrounded by dystrophic neuritis, they are called tangle associated neuritic clusters (527). Although NFTs correlate better with dementia severity than amyloid plaques (490), they can be absent in the neocortex in 10% of patients with AD and in as many as the 50% of mild AD cases (528).

Neuropil threads are most commonly seen in AD and only rarely identified in other tauopathies such as corticobasal degeneration (529). They are short tortuous neuronal dendrites filled with abnormal tau (527). Dystrophic neurites are tau-containing dendritic structures that are seen in the periphery of the senile plaques.

Post-translational phosphorylation of tau is an additional source of microheterogeneity (530). During brain development, tau is phosphorylated at many residues with GSK-3β, cdk 5, and MAPK (531). In vitro, tau can be phosphorylated on multiple sites by several kinases, too (for a review, see (532)). Most of the in vitro phosphorylation sites are located within the microtubule interacting region (repeat domain) and sequences flanking the repeat domain. Many of these sites are also phosphorylated in PHF-tau (533, 534). In fact, 10 major phosphorylation sites have been identified in tau isolated from PHFs from patients with AD (533, 534). Hyperphosphorylation was shown to be accompanied by the transformation from the unfolded state of tau into a partially folded conformation (535, 536), accelerating the self-assembly of this protein into paired helical filaments in vitro (523). To analyze the potential role of tau hyperphosphorylation in tauopathies, mutated tau proteins have been produced, in which all 10 serine/threonine residues known to be highly phosphorylated in PHF-tau were substituted for negatively charged residues, thus producing a model for a defined and permanent hyperphosphorylation-like state of tau protein (537). It has been demonstrated that, like hyperphosphorylation, glutamate substitutions induce compact structure elements and SDS-resistant conformational domains in tau protein, as well as lead to the dramatic acceleration of its fibrillation (537).

Prior the aggregation, tau protein was shown to be in a mostly random coil-like state. This conclusion followed from the conformational analysis of this protein by CD, Fourier transform infrared spectroscopy, small angle X-ray scattering and biochemical assays (538). Analysis of the primary structure reveals a very low content of hydrophobic amino acids and a high content of charged residues, which was sufficient to explain the lack of folding (538). Analysis of the hydrodynamic radii confirms a mostly disordered structure of various tau isoforms and tau domains. However, the protein was further unfolded in the presence of high concentrations of strong denaturant GdmCl, indicating the presence of some residual structure.
Protein intrinsic disorder in degenerative diseases

This conclusion was supported by a FRET-based approach where the distances between different domains of tau were determined. The combined data show that tau is mostly disordered and flexible but tends to assume a hairpin-like overall fold which may be important in the transition to a pathological aggregate (538).

Intriguingly, purified recombinant tau isoforms do not detectably aggregate over days of incubation under physiological conditions. However, aggregation and fibrillation can be dramatically accelerated by the addition of anionic surfactants (539). Based on the detailed analysis of tau fibrillation in the presence of anionic inducers using a set of spectroscopic techniques (CD and reactivity with thioflavin S and ANS fluorescent probes) it has been established that the inducer stabilized a monomeric partially folded species with the structural characteristics of a pre-molten globule state (540). The stabilization of this intermediate was sufficient to trigger the fibrillation of full-length tau protein (540).

5.3. Prion protein and prion diseases

Prion diseases are a group of incurable, fatal neurodegenerative maladies that afflict mammals. These diseases, collectively referred to as the transmissible spongiform encephalopathies (TSEs), are caused by the pathological deposition of the prion protein (PrP) in its aggregated form. TSEs include Creutzfeldt-Jakob disease (CJD), Gerstmann-Sträussler-Scheinker (GSS) disease, fatal familial insomnia (FFI) and kuru in humans, scrapie in sheep, bovine spongiform encephalopathy (BSE) in cattle, and chronic wasting disease (CWD) in mule deer and elk (541). The most important aspect is the transmission of PrP aggregates from one individual or species to another, causing prion diseases. Prion diseases are unique among all illnesses in that they can manifest as sporadic, genetic or infectious maladies. Similar to many other neurodegenerative diseases, the sporadic form of prion disease accounts for ~80-90% of cases whereas the genetic forms account for 10 to 20% (542, 543). Infection by exogenous prions seems to be responsible for <1% of all human cases of prion disease (544).

The characteristic pathological features of TSEs are spongiform degeneration of the brain and accumulation of the abnormal, protease-resistant PrP isoform in the central nervous system, which sometimes forms amyloid-like plaques. The prion concept was introduced in 1982 in order to explain a vast body of scientific data, much of which argued the pathogen causing scrapie is devoid of nucleic acid but contains a protein that is essential for infectivity (545). Prions are unprecedented infectious pathogens that give rise to invariably fatal neurodegenerative diseases via an entirely novel mechanism of disease.

Native prion protein (PrP) is attached to the extracellular plasma membrane surface by a glycosylphosphatidylinositol lipid anchor and undergoes endocytosis. The N-terminal region of about 100 amino acids in PrP (from amino acid 23 to 126) is largely unstructured in the isolated molecule in solution (546). The C-terminal domain is folded into a largely α-helical conformation (three α-helices and a short antiparallel β-sheet) and stabilized by a single disulfide bond linking helices 2 and 3 (547). The central event in the pathogenesis of prion diseases is believed to be a major conformational change of the C-terminal region of the PrP from an α-helical (PrP\(\alpha\)) to a β-sheet-rich isoform (PrP\(\beta\)), and PrP\(\beta\) propagates itself by causing the conversion of PrP\(\alpha\) to PrP\(\beta\). Although unstructured in the isolated molecule, the N-terminal region contains tight binding sites for Cu\(^{2+}\) ions and acquires structure following copper binding (548, 549).

Two pathological GSS-like mutations, Y145Stop and Q160Stop, result in C-terminal truncated isoforms. The truncation occurs just after the central region from amino acid 90 to 145, which was shown to be converted into β-sheet as a result of the PrP\(\beta\) to PrP\(\beta\) conversion (550, 551). Structural properties and aggregation propensities of these variants in vitro were analyzed by a variety of biophysical techniques (552). It has been shown that although both proteins are substantially disordered, a continuous stretch of positive secondary chemical NMR shifts was found for residues 144-154 in Q160Stop protein, indicative of helical structure. This clearly demonstrated that although the vast majority of a polypeptide chain is substantially disordered, a significantly populated helix 1 is present in human Q160Stop protein (552). Q160Stop protein was shown to fibrillate faster than shorter Y145Stop variant. Intriguingly, helix 1 was not converted to the β-sheet during the protein aggregation. Based on the results of this analysis it has been concluded that the highly charged helix 1 is involved in the aggregation of Q160Stop protein likely via the formation of intermolecular salt bridges (552).

Investigations of the steps required for prion propagation and neurodegeneration in transgenic mice expressing chimeric mouse–hamster–mouse or mouse–human–mouse PrP transgensics indicated that the last 50 residues in the disordered N-terminal region play a particularly important role in the interaction of PrP\(\beta\) with PrP\(\beta\) leading to the conversion of the former to the latter (553, 554). Those residues are largely unordered or weakly helical in the full-length PrP\(\beta\) (555, 556), but are predicted to be β-structure in PrP\(\beta\) (545). These observations emphasize a crucial role of the disordered N-terminal region in the modulation of PrP aggregation. Several kinetics studies have revealed the existence of partially folded intermediates for the PrP (545, 557, 558), and it is reasonable to assume that fibrillation requires partial unfolding of the C-terminal domain prior to self-association.

5.4. α-Synuclein and synucleinopathies

Synucleinopathies (see Table 2) is a group of neurodegenerative disorders characterized by fibrillar aggregates of α-synuclein protein in the cytoplasm of selective populations of neurons and glia (559-562). Clinically, synucleinopathies are characterized by a chronic and progressive decline in motor, cognitive, behavioral, and autonomic functions, depending on the distribution of the lesions. Because of clinical overlap, differential diagnosis is sometimes very difficult (563). Depending on the type of
Table 2. Human neurodegenerative disorders with α-synuclein deposits

<table>
<thead>
<tr>
<th>Diseases with neuronal inclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal aging</td>
</tr>
<tr>
<td>Parkinson's disease</td>
</tr>
<tr>
<td>- Idiopathic</td>
</tr>
<tr>
<td>- Neurotoxicant-induced (incidental)</td>
</tr>
<tr>
<td>- Familial</td>
</tr>
<tr>
<td>- With α-synuclein point mutations</td>
</tr>
<tr>
<td>- With α-synuclein gene triplication</td>
</tr>
<tr>
<td>- With mutations in other proteins</td>
</tr>
<tr>
<td>- Pure autonomic failure</td>
</tr>
<tr>
<td>- Lewy body dysphagia</td>
</tr>
<tr>
<td>Parkinsonism plus syndromes</td>
</tr>
<tr>
<td>- Sporadic</td>
</tr>
<tr>
<td>- Progressive supranuclear palsy</td>
</tr>
<tr>
<td>- Olivoponto cerebellar atrophy (Shy-Drager syndrome)</td>
</tr>
<tr>
<td>- Cortical-basal ganglionic degeneration</td>
</tr>
<tr>
<td>- Sporadic pallidal degeneration</td>
</tr>
<tr>
<td>- Bilateral striatopallidal dentate calcinosis</td>
</tr>
<tr>
<td>- Parkinsonism with neuroacanthocytosis</td>
</tr>
<tr>
<td>- Familial</td>
</tr>
<tr>
<td>- Familial diffuse Lewy body disease</td>
</tr>
<tr>
<td>- Familial dementia with swollen achromatic neurons and cortico-basal inclusion bodies</td>
</tr>
<tr>
<td>- Frontotemporal dementia with parkinsonism linked to chromosome 17</td>
</tr>
<tr>
<td>- Associated with psychiatric disturbances</td>
</tr>
<tr>
<td>- Associated with respiratory disturbances</td>
</tr>
<tr>
<td>- Associated with dystonia</td>
</tr>
<tr>
<td>- Associated with myoclonus and seizures</td>
</tr>
<tr>
<td>- Familial progressive supranuclear palsy</td>
</tr>
<tr>
<td>Alzheimer's disease</td>
</tr>
<tr>
<td>- Sporadic</td>
</tr>
<tr>
<td>- Familial with APP mutation</td>
</tr>
<tr>
<td>- Familial with PS-1 mutation</td>
</tr>
<tr>
<td>- Familial with other mutations</td>
</tr>
<tr>
<td>- Familial British dementia</td>
</tr>
<tr>
<td>Lewy body diseases</td>
</tr>
<tr>
<td>- Dementia with Lewy bodies</td>
</tr>
<tr>
<td>- Pure form - transitional/limbic</td>
</tr>
<tr>
<td>- Pure form - neocortical</td>
</tr>
<tr>
<td>- Diffuse Lewy body disease</td>
</tr>
<tr>
<td>- Common form</td>
</tr>
<tr>
<td>- Pure form</td>
</tr>
<tr>
<td>- Lewy body variant of Alzheimer’s disease</td>
</tr>
<tr>
<td>- Incidental Lewy body disease</td>
</tr>
<tr>
<td>- Lewy body dementia</td>
</tr>
<tr>
<td>- Senile dementia of Lewy body type</td>
</tr>
<tr>
<td>- Dementia associated with cortical Lewy bodies</td>
</tr>
<tr>
<td>Down’s syndrome</td>
</tr>
<tr>
<td>Amyotrophic lateral sclerosis-parkinsonism/dementia complex of Guam</td>
</tr>
<tr>
<td>Neuroaxonal dystrophies</td>
</tr>
<tr>
<td>- Neurodegeneration with brain iron accumulation, type 1 (Hallervorden-Spatz syndrome or adult neuroaxonal dystrophy)</td>
</tr>
<tr>
<td>- Motor neuron disease</td>
</tr>
<tr>
<td>Amyotrophic lateral sclerosis</td>
</tr>
<tr>
<td>- Familial</td>
</tr>
<tr>
<td>- Sporadic</td>
</tr>
<tr>
<td>Tauopathies</td>
</tr>
<tr>
<td>- Frontotemporal degeneration/dementia</td>
</tr>
<tr>
<td>- Pick’s disease</td>
</tr>
<tr>
<td>- Post-encephalitic parkinsonism</td>
</tr>
<tr>
<td>- Dementia pugilistica</td>
</tr>
<tr>
<td>- Amyotrophic grain disease</td>
</tr>
<tr>
<td>- Corticobasal degeneration</td>
</tr>
<tr>
<td>Prion diseases</td>
</tr>
<tr>
<td>- Transmissible spongiform encephalopathies</td>
</tr>
<tr>
<td>- Sporadic</td>
</tr>
<tr>
<td>Creutzfeldt-Jakob disease</td>
</tr>
<tr>
<td>- Familial</td>
</tr>
<tr>
<td>- Familial Creutzfeldt-Jakob disease</td>
</tr>
<tr>
<td>- Gertsmann-Strausser-Scheinker syndrome</td>
</tr>
<tr>
<td>- Infectious</td>
</tr>
<tr>
<td>- Iatrogenic Creutzfeldt-Jakob disease</td>
</tr>
<tr>
<td>- Variant Creutzfeldt-Jakob disease</td>
</tr>
<tr>
<td>- Kuru</td>
</tr>
</tbody>
</table>
Protein intrinsic disorder in degenerative diseases

- Fatal familial insomnma
- Ataxia telangiectatica
- Meige's syndrome

Diseases with neuronal and glial inclusions
- Multiple system atrophy
- Shy-Drager syndrome
- Striatonigral degeneration (MSA-P)
- Olivopontocerebellar atrophy (MSA-C)

Pathology, α-synuclein inclusions are present in neurons (both dopaminergic and non-dopaminergic), where they can be deposited in perikarya or in axonal processes of neurons, and in glia. At least five morphologically different α-synuclein containing inclusions have been determined: LBs, LNs (dystrophic neurites), glial cytoplasmic inclusions (GCIs), neuronal cytoplasmic inclusions and axonal spheroids. Some of the disorders associated with the α-synuclein depositions are discussed below to illustrate a wide range of pathological manifestations in synucleinopathies.

5.4.1. α-Synuclein and Parkinson's disease

PD is the most common aging-related movement disorder and second most common neurodegenerative disorder after AD. It is estimated that ~1.5 million Americans are affected by PD. Since only a small percentage of patients are diagnosed before the age of 50, PD is generally considered as an aging-related disease, and approximately one of every 100 persons over the age of 55 in the US suffers from this disorder (564). PD is a slowly progressive disease that affects neurons of the substantia nigra, a small area of cells in the mid-brain. Gradual degeneration of the dopaminergic neurons causes a reduction in the dopamine content. This, in turn, can produce one or more of the classic signs of PD: resting tremor on one (or both) side(s) of the body; generalized slowness of movement (bradykinesia); stiffness of limbs (rigidity); and gait or balance problems (postural dysfunction). The substantia nigra consists of ~400,000 nerve cells, which begin to pigment after birth and are fully pigmented at age 18. The symptoms of PD become apparent after more than ~70% dopaminergic neurons die. This neurodegeneration is characterized by the dramatic depigmentation of the substantia nigra, indicating that there is a relationship between pigmentation and function of the substantia nigra. The “normal” rate of nigral cell loss is ~2,400 per a year. Thus, if an unaffected person lives to be 100 years old he (she) will probably develop PD. In PD, the neuron loss is accelerated. Although, it is unknown why nerve cells loss accelerates, it appears to be due to a combination of genetic susceptibility and environmental factors. Some surviving nigral dopaminergic neurons contain cytosolic filamentous inclusions known as LBs when found in the neuronal cell body, or LNs when found in axons (565, 566).

Several observations implicate α-synuclein in the pathogenesis of PD. Autosomal dominant early-onset PD was shown to be induced in a small number of kindreds as a result of three different missense mutations in the α-synuclein gene, corresponding to A30P, E46K, and A53T substitutions in α-synuclein (567-569) or as a result of the hyper-expression of the wild type α-synuclein protein due to gene triplication (570-572). Antibodies to α-synuclein detect this protein in LBs and LNs. A substantial portion of fibrillar material in these specific inclusions was shown to be composed of α-synuclein, and insoluble α-synuclein filaments were recovered from purified LBs (573, 574). The production of wild type α-synuclein in transgenic mice (575) or of WT, A30P, and A53T in transgenic flies (576), leads to motor deficits and neuronal inclusions reminiscent of PD. Under the particular conditions, cells transfected with -synuclein might develop LB-like inclusions. Other important observations correlating α-synuclein and PD pathogenesis were reviewed in more detail elsewhere (367, 377, 561, 577-579).

5.4.2. α-Synuclein in dementia with Lewy bodies and other Lewy body disorders

5.4.2.1. Dementia with Lewy Bodies.

Dementia with Lewy bodies (DLB), being the second most frequent neurodegenerative dementing disorder after AD, is a common form of late-onset dementia that exists in a pure form or overlaps with the neuropathological features of AD. This disease is characterized clinically by neuropsychiatric changes often with marked fluctuations in cognition and attention, hallucinations, and parkinsonism (580). Similar to PD, neuropathological hallmarks of DLB are numerous LBs and LNs in the substantia nigra, which are strongly immunoreactive for -synuclein (573). However, unlike PD, DLB is characterized by large numbers of LBs and LNs in cortical brain areas (581). It has been noted that filaments from LBs in DLB are decorated by -synuclein antibodies (574, 582, 583), and that their morphology closely resembles that of filaments extracted from the substantia nigra of PD brains (574, 583). DLB and PD with dementia, being different in the temporal course of the disease, share most of the same clinical and neuropathological features and are often considered as belonging to a spectrum of the same disease (584-586). It is well recognized now that the incidence of dementia in PD is higher than expected from aging alone (580), as dementia affects about 40% of PD patients (587), and the incidence of dementia in PD patients is up to six times greater than observed in normal aged matched control subjects (588).

5.4.2.3. Amyotrophic lateral sclerosis-parkinsonism/dementia complex of Guam.

Guam disease is another example of PD and dementia junction. Guam disease is a neurodegenerative disorder with unusually high incidence among the Chamorro people of Guam (589-591). The neurotoxic plant Cycas circinalis, a traditional source of food and medicine used by the Chamorro people, plays a role in the development of Guam ALS-parkinsonism-dementia (591). Intriguingly, recent studies revealed that in general three
neurodegenerative disorders, ALS, dementia, and PD, co-occur within families more often than expected by chance, suggesting that there may be a shared genetic susceptibility to these disorders (592).

5.4.2.3. Other Lewy body diseases

Several peripheral and central areas of the nervous system can be affected by the LB deposition. Besides already discussed substantia nigra, this includes hypothalamic nuclei, nucleus basalis of Meynert, dorsal raphe, locus ceruleus, dorsal vagus nucleus, and intermediolateral nucleus (593). A 'neuritic' form of LB was also described in the dorsal vagus nucleus, sympathetic ganglia, and in intramural autonomic ganglia of the gastrointestinal tract, as well cases were demonstrated with extensive cortical and basal ganglia involvement (581, 594). This broad spectrum of the nervous system regions potentially affected by LB formation produces great variability in the disease manifestation and LB pathology is also a characteristic feature of several rarer diseases, such as pure autonomic failure, LB dysphagia, incidental LB disease (578, 579). Pure autonomic failure (also known as Bradbury-Eggleston syndrome) (595) and LB dysphagia (596) are the results of the predominant involvement of the peripheral nervous system with minimal central nervous system involvement. In incidental LB disease, ~5%-10% of asymptomatic individuals have insignificant numbers of LBs bodies, usually located in substantia nigra (597).

5.4.3. α-Synuclein and Alzheimer's disease

Detailed analysis of the α-synuclein immunoreactivity in the brains from the patients with sporadic AD revealed the presence of α-synuclein-positive inclusions resembling LBs and LNs in ~50% cases studied (598). α-Synuclein-positive LB-like intra-cytoplasmic inclusions were found in the amygdala, the temporal cortex, the parahippocampal gyrus, and in the parietal cortex, whereas LN-like inclusions were abundant in the amygdala, the CA2/3 region of hippocampus formation, parahippocampal gyrus, the temporal cortex, substantia nigra, locus ceruleus, the frontal cortex, and in the parietal cortex (598).

5.4.4. α-Synuclein and Down’s Syndrome

Down’s syndrome is a genetic disorder characterized by an extra chromosome 21 (trisomy 21, i.e., instead of having the normal 2 copies of chromosome 21, the Down’s syndrome patient has 3 copies of this chromosome). The person with Down’s syndrome has mild mental retardation, short stature, a flattened facial profile, a risk of multiple malformations (including heart malformations; duodenal atresia, where part of the small intestines is not developed and leukemia), and susceptibility to early-onset AD. Incidence of this disorder among the newborn is estimated at 0.03%, whereas in the general population it is approximately 0.01%. The difference reflects the early mortality. The analysis of Down’s syndrome with Alzheimer pathology revealed presence of numerous LBs and LNs in the neurons of the limbic areas, predominantly of the amygdala. Similar lesions were less common in other regions of these brains (599, 600). Importantly, in the vast majority of cases examined no LBs and LNs were detected in the substantia nigra and locus ceruleus, and there was no significant neuronal loss in the substantia nigra.

5.4.5. α-Synuclein and Multiple System Atrophy

Multiple system atrophy (MSA) is an adult-onset progressive neurodegenerative disorder of unknown etiology which is characterized clinically by any combination of parkinsonian, autonomic, cerebellar or pyramidal symptoms and signs, and pathologically by cell loss, gliosis and GCIs in several brain and spinal cord structures. Most patients affected by MSA deteriorate rapidly and survival beyond ten years after disease onset is unusual. It is believed that the motor impairment in MSA results from L-DOPA-unresponsive parkinsonism, cerebellar ataxia and pyramidal signs, with 80% of MSA cases showing predominant parkinsonism (MSA-P) due to underlying striatoni gral degeneration, and the remaining 20% developing predominant cerebellar ataxia (MSA-C) associated with olivopontocerebellar atrophy (601).

Autonomic dysfunction including urogenital failure and orthostatic hypotension is common in both motor presentations, MSA-P and MSA-C, reflecting degenerative lesions of central autonomic pathways (602). Distinguishing MSA-P from PD is problematic at early stages owing to PD-like features in MSA-P, including a transient L-dopa response in some patients (603). MSA is less common than PD as epidemiological studies suggested a prevalence of 1.9-4.9 people per 100,000 and an incidence of 3 patients per 100,000 people per year (604-606). Histologically, MSA is characterized by the variable neuron loss in the striatum, substantia nigra pars compacta, cerebellum, pons, inferior olives and intermediolateral column of the spinal cord (607). The histological hallmark of MSA is the presence of argyrophilic fibrillary inclusions in the oligodendrocytes, referred to as GCIs, which are also known as Papp-Lantos bodies (608). Fibrillar inclusions are also found in the neuronal somata, axons, and nucleus. Neuronal cytoplasmic inclusions are frequently found in the pontine and inferior olivary nuclei (609). It has been established that α-synuclein is a major component of glial and neuronal inclusions in MSA (583, 609). Although both LBs and GCIs contain α-synuclein, they are differently localized, with α-synuclein inclusions being neuronal in PD and DLB, and oligodendrogial in MSA. This suggests the existence of a unique pathogenic mechanism that ultimately lead to neuron loss via disturbance of axonal function (608). In MSA, besides formation of GCIs α-synuclein also aggregates in the cytoplasm, axons and nuclei of neurons, and the nuclei of oligodendroglia. The relationship between GCIs and these additional α-synuclein deposition sites is not understood (608).

5.4.6. α-Synuclein and Neurodegeneration with Brain Iron Accumulation Type 1 (NBIA1)

Neurodegeneration with brain iron accumulation type 1 (NBIA1) (formerly known as Hallervorden-Spatz disease (HSD) or adult neuroaxonal dystrophy) represents a rare progressive neurodegenerative disorder that occurs in both sporadic as well as in familial forms. Clinically, NBIA1 is characterized by rigidity, dystonia, dyskinesia,
Protein intrinsic disorder in degenerative diseases

and choreoatethosis (610-613), together with dysarthria, dysphagia, ataxia, and dementia (613-615). Symptoms usually present in late adolescence or early adult life and this disease is persistently progressive (610, 614, 615). The histopathological hallmarks of NBIA1 include neuronal loss, neuraxonal spheroids, and iron deposition in the globus pallidus and substantia nigra pars compacta, as well as by the presence of the LB-like and GCI-like inclusions and dystrophic neuritis (614). NBIA1 is characterized by an association of extrapyramidal movement disorders with neuroaxonal dystrophy (NAD) and iron accumulation in the basal ganglia. It represents a pantothenate kinase-associated neurodegeneration caused by the PANK2 gene linked to chromosome 20p12.3-13 (616). It has been shown that the LB-like inclusions throughout the cortex and brainstem, axonal swellings, and rare GCI-like inclusions of the midbrain clearly possess α-synuclein immunoreactivity (617-619). Importantly, axonal spheroids were also shown to contain α-synuclein (619, 620).

5.5. β- and γ-Synucleins in Parkinson’s disease and dementia with Lewy bodies

Synucleins are members of a family of closely related presynaptic proteins that arise from three distinct genes, described currently only in vertebrates (621). This family includes: α-synuclein, which also known as the non-amyloid component precursor protein (NACP) or synelfin (364, 622, 623); β-synuclein, also referred to as phosphoneuro-protein 14 or PNP14 (623-625) and γ-synuclein, also known as breast cancer-specific gene 1 or BCSG1 and persyn (626-629).

Human β-synuclein is a 134-aa neuronal protein showing 78% identity to α-synuclein. The α- and β-synucleins share a conserved C-terminus with three identically placed tyrosine residues. However, β-synuclein is missing 11 residues within the specific non-amyloid component (NAC) region (630, 631). The activity of β-synuclein may be regulated by phosphorylation (624). This protein, like α-synuclein, is expressed predominantly in the brain, however, in contrast to α-synuclein, β-synuclein is distributed more uniformly throughout the brain (632, 633). Besides the central nervous system β-synuclein was also found in Sertoli cells of the testis (634, 635), whereas α-synuclein was found in platelets (636).

The third member of the human synuclein family is the 127-aa γ-synuclein, which shares 60% similarity with α-synuclein at the amino acid sequence level (630, 631). This protein specifically lacks the tyrosine rich C-terminal signature of α- and β-synucleins (631). γ-Synuclein is abundant in spinal cord and sensory ganglia (628). Interestingly, this protein is more widely distributed within the neuronal cytoplasm than α- and β-synucleins, being present throughout the cell body and axons (628). It was also found in metastatic breast cancer tissue (627) and epidermis (637).

It has recently been established that in addition to the traditional α-synuclein-containing LBs and LNs, the development of PD and DLB is accompanied by appearance of novel α-, β- and γ-synuclein-positive lesions at the axon terminals of hippocampus (638). These pathological vesicular-like lesions located at the presynaptic axon terminals in the hippocampal dentate, hilar, and CA2/3 regions have been co-stained by antibodies to α- and β-synucleins, whereas antibodies to γ-synuclein detect previously unrecognized axonal spheroid-like inclusions in the hippocampal dentate molecular layer (638). This broadens the concept of neurodegenerative “synucleinopathies” by implicating β- and γ-synucleins, in addition to α-synuclein, in the onset/progression of these two diseases.

Structural properties of the members of synuclein family have been compared using several physico-chemical methods (639). All three proteins showed far-UV CD spectra typical of an unfolded polypeptide chain. Interestingly, α- and γ-synucleins possessed almost indistinguishable spectra, whereas the far UV-CD spectrum of β-synuclein showed a slightly increased degree of disorder. The increased unfoldedness of β-synuclein was further confirmed by hydrodynamic studies performed by size-exclusion chromatography and SAXS (639). This emphasized the importance of the NAC region to maintain the residual partially collapsed structure in α- and γ-synucleins.

Conformational analysis revealed that α-, β-, and γ-synucleins are typical natively unfolded proteins that are able to adopt comparable partially folded conformations at acidic pH or at high temperature (639). Although both α- and γ-synucleins were shown to form fibrils, β-synuclein did not fibrillate, being incubated under the same conditions (639). However, even non-amyloidogenic β-synuclein can be forced to fibrillate in the presence of some metals (Zn2+, Pb2+, and Cu2+) (640).

Intriguingly, the addition of either β- or γ-synuclein in a 1:1 molar ratio to α-synuclein solution substantially increased the duration of the lag-time and dramatically reduced the elongation rate of α-synuclein fibrillation (639). Fibrillation was completely inhibited at a 4:1 molar excess of β- or γ-synuclein over α-synuclein (639). β-Synuclein inhibited α-synuclein aggregation in an animal model, too (641). This suggests that β- and γ-synucleins may act as regulators of α-synuclein fibrillation in vivo, potentially acting as chaperones. Therefore, one possible factor in the etiology of PD would be a decrease in the levels of β- or γ-synucleins (639). The logical question is how the chaperone roles of β- and γ-synucleins are compatible with their ability to fibrillate on their own. The answer to this question is in the specific details of the fibrillation kinetics of three synucleins: conditions promoting β-synuclein aggregation were very different from conditions favoring α-synuclein fibrillation (640), whereas γ-synuclein fibrillated slower than α-synuclein (639). The ability to form amyloid fibrils was recently shown for a typical member of the molecular chaperone family, co-chaperonin GroES (3).
5.6. Polyglutamine repeat diseases associated with aggregation of huntingtin, ataxins, androgen receptor, and atrophin-1

5.6.1. Polyglutamine repeat diseases

Currently, there are at least nine known hereditary diseases in which the expansion of a CAG repeat in the gene leads to neurodegeneration (642, 643). Table 1 shows that these polyglutamine repeat diseases includes HD, Kennedy disease (also known as spinal and bulbar muscular atrophy, SBMA), spinocerebellar ataxia type 1 (SCA1), dentatorubral-pallidolysian atrophy (DRPLA), spinocerebellar ataxia type 2 (SCA2), Machado-Joseph disease (MJD/SCA3), SCA6, SCA7 and SCA17. These diseases are accompanied by the progressive death of neurons, with insoluble, granular, and fibrous deposits being found in the cell nuclei of the affected neurons. The neurotoxicity in these diseases is due to the expansion of the (CAG)n-encoded polyglutamine (polyQ) repeat, which leads to the formation of amyloid fibrils and neuronal death. In HD, the CAG repeat that encodes the polyQ region is part of exon 1 in the 3,140-residue huntingtin protein (644). The polyQ repeat varies between 16 and 37 residues in healthy individuals, and who are afflicted by disease have repeats of >38 residues.

The age of onset and the severity of the progression of SCA1, an autosomal-dominant neurodegenerative disorder characterized by ataxia and progressive motor deterioration, are directly correlated with the length of the polyQ segment in ataxin-1, a nuclear protein of ~800 residues (645-647). When the number of glutamine residues in the polyQ tract exceeds a threshold (39-44 glutamine residues), ataxin-1 aggregates with granular or fibrillar morphologies accumulate intranuclearly and eventually lead to cell death (648, 649).

SBMA is linked to the expansion of a Q-rich segment in the androgen receptor (650); healthy individuals have a 15- to 31-residue polyQ segment, and individuals who are afflicted with the disease have 40-62 Q residues. Intriguingly, in the human androgen receptor there are three polyglutamine repeats ranging in size from five to 22 residues, stretches of seven prolines and five alanines, and a polyglycine repeat of 24 residues. Polymorphisms in the length of the largest polyglutamine and the polyglycine repeats of the androgen receptor have been associated with a number of clinical disorders, including prostate cancer, benign prostatic hyperplasia, male infertility and rheumatoid arthritis (651).

The onset of the DRPLA, another progressive neurodegenerative disorder characterized by a distinctive pathology in the cerebellar and pallidal outflow pathways, is inversely correlated with the polyQ repeat size in the corresponding DRPLA protein (also known as atrophin-1), a product of the gene on chromosome 12p (652). The repeat size varied from 7-23 in normal individuals and was expanded to 49-75 in DRPLA patients.

5.6.2. Huntingtin and structure of polyglutamine stretches

Huntingtin, a protein with an estimated molecular mass of 350 kDa, contains a polyglutamine tract near its N terminus that when expanded beyond 37 glutamines causes HD (644). The N terminus of wild-type huntingtin interacts with proteins involved in nuclear functions, including HYP/A/FBP-11, which functions in pre-mRNA processing (splicesome function) (653), nuclear receptor co-repressor protein (NCoR) (654), which plays a role in the repression of gene activity, and p53 (655), a tumor suppressor involved in regulation of the cell cycle. Full-length huntingtin contains candidate binding sites for other proteins with nuclear functions. Huntingtin contains a PDLS motif, a candidate-binding site for the transcriptional corepressor C-terminal binding protein (CtBP) (656), suggesting that huntingtin may play a role in transcriptional repression.

The localization and potential function of normal and mutated huntingtin in the nucleus was suggested to be important for understanding HD pathogenesis. For example, N-terminally mutated huntingtin was shown to be toxic when targeted to the nucleus of cultured striatal neurons (657). Mutated huntingtin has been implicated in abnormal transcriptional repression in HD. In cellular systems, short N-terminal mutated huntingtin fragments disrupt transcriptional regulation, which occurs through a mechanism involving sequestration of transcription factors including p53 (655), TATA-box-binding protein (TBP) (658), and CREB-binding protein (659) into huntingtin-positive aggregates. These results suggest that the N terminus of mutated huntingtin may disrupt neuronal function in HD by interfering with nuclear organization and transcriptional regulation. Full-length huntingtin was shown to co-immunoprecipitate with the transcriptional corepressor C-terminal binding protein, and polyglutamine expansion in huntingtin reduced this interaction (660). Interestingly, although full-length wild-type and mutated huntingtin both repressed transcription when targeted to DNA, N-terminally truncated huntingtin was shown to repress transcription, whereas the corresponding wild-type fragment did not (660).

Proteolytic cleavage of mutated huntingtin is suggested to play a key role in the pathogenesis of HD. Huntingtin was shown to be cleaved by caspases and calpains within a region between 460-600 amino acids from the N-terminus. Furthermore, two smaller N-terminal fragments produced by unknown protease have been described as cp-A and cp-B (661). In fact, based on the analysis of human HD patients, animal models and cell-based models of HD it has been suggested that truncated polyglutamine-containing fragments are more toxic than full-length huntingtin (662).

The mechanistic hypothesis linking CAG repeat expansion to toxicity involves the tendency of longer polyQ sequences, regardless of protein context, to form insoluble aggregates (369, 663-670). To help evaluate various possible mechanisms, the biophysical properties of a series of simple polyglutamine peptides have been analyzed. The far-UV CD spectra of polyQ peptides with repeat lengths of 5, 15, 28 and 44 residues were shown to be nearly identical and were consistent with a high degree of random coil structure, suggesting that the length-dependence of disease is not related to a conformational change in the monomeric
states of expanded polyQ sequences (668). In contrast, there was a dramatic acceleration in the spontaneous formation of ordered, amyloid-like aggregates for polyQ peptides with repeat lengths of greater than 37 residues. Several studies established the role of partially folded intermediates of polyglutamine-repeat proteins as key species in fibrillation (669, 671, 672).

Huntingtin was shown to interact with more than 200 proteins (673). One of these huntingtin interactors, huntingtin yeast-two hybrid protein K (HYPK) was recently identified as a typical IDP using a set of biophysical and biochemical techniques (673). Among the experimental data supporting this conclusion there were aberrant electrophoretic mobility [the molecular weight of HYPK determined by gel electrophoresis was found to be about 1.3-folds (~22 kDa) higher than that obtained from mass spectrometric analysis (16.9 kDa)]; increased hydrodynamic dimensions [in size exclusion chromatography experiment, HYPK was eluted as a protein with the hydrodynamic radius which was ~1.5-folds (23 Å) higher than that expected for globular proteins of equivalent mass (17.3 Å)]; random coil characteristics of far-UV CD spectra; and highly sensitive to limited proteolysis by trypsin and papain (673). Subsequent analysis of HYPK revealed that this huntingtin interacting protein was able to reduce aggregates and apoptosis induced by N-terminal huntingtin with 40 glutamines in Neuro2a cells and exhibited chaperone-like activity (674).

5.6.3. Dentatorubral-pallidolusian atrophy protein (atrophin-1)

Investigations of the DRPLA gene (encoding for atrophin-1) indicate that it is widely expressed in brain and other tissues as a 4.5-kb transcript with an open reading frame encoding 1184 amino acids (675-677). The rat atrophin-1 coding sequence is 88% identical to the coding sequence of human atrophin-1 at the level of DNA and 94% identical at the protein level, but encodes a shorter glutamine repeat that is followed by a series of alternating glutamine and proline residues (678, 679). The predicted molecular mass of the atrophin-1 gene product is 124 kDa, yet atrophin-1 appeared to migrate at about 200 kDa (680).

5.6.4. Androgen receptor

CD analysis of a region of the androgen receptor N-terminal domain lacking the largest polyglutamine stretch, but containing the remaining repeats, showed that it lacked stable tertiary structure in aqueous solutions (651). Detailed conformational studies using a combination of experimental and computational techniques revealed that the AF1 transactivation domain is in the molten globule-like conformation (681). In fact, this region of the receptor was predicted to contain long disordered regions, when analyzed by amino acid composition, PONDR®-based predictions for the androgen receptor AF1 suggests that this domain possesses properties consistent with a dynamic conformation and to fall into a “collapsed disorder class” of proteins, typical of the molten globule folding intermediate (276, 328). This conclusion was confirmed by the analysis of a hydrophobic fluorescence probe, ANS, binding and by size-exclusion chromatography (681). The results of this analysis suggest that native androgen receptor AF1 exists in a collapsed disordered conformation, distinct from extended disordered (random coil) and a stable globular fold (681).

5.6.5. Ataxin-2

SCA2 is an autosomal-dominantly inherited, neurodegenerative disorder, caused by the expansion of an unstable CAG/polyQ repeat located at the N-terminus of ataxin-2 protein. The age of onset of SCA2 is in the third to fourth decade. The characteristic phenotypic features of SCA2 are the degeneration of specific vulnerable neuron populations and the presence of intracellular aggregations of the mutated protein in affected neurons. Ataxin-2 has 1312 residues (including 22 glutamines of the polyQ stretch) and a molecular mass of ~140 kDa. Ataxin-2 is a highly basic protein except for one acidic region (amino acid 254-475) containing 46 acidic amino acids (682). This region consists of two predicted globular domains, Lsm (Like Sm, amino acid 254-345) and LsmAD (Lsm-associated domain, amino acid 353-475). The LsmAD domain contains a clathrin-mediated trans-Golgi signal (YDS, amino acid 414-416) and an endoplasmic reticulum (ER) exit signal (ERD, amino acid 426-428). This domain is composed mainly of u-helices according to the results from secondary structure prediction servers. The rest of ataxin-2 outside of the Lsm and LsmAD domains is only weakly conserved in eukaryotic ataxin-2 homologues and is predicted to be intrinsically disordered (682).

5.6.6. Ataxin-3

Human ataxin-3, the protein related to SCA3/MJD, is a ubiquitously expressed 41 kDa protein whose polyQ tract contains 12-40 glutamines in normal individuals and 55-84 glutamines in the pathogenic form (642, 643). Ataxin-3 is present in the genomes of several species, from nematodes to human, including plants (683). Alignment of the ataxin-3 family shows a conserved N-terminal block that corresponds to the sequence motif named Josephin (residues 1-198 in the human protein) (683). The C-terminus is non-conserved throughout different species and contains long stretches of low complexity regions which include the polyQ tract, preceded by a highly charged region (683).

Human ataxin-3 was analyzed by a range of biophysical and biochemical techniques, including limited proteolysis, CD and NMR spectroscopies (684). The deconvolution of the far-UV CD spectra indicated that ataxin-3 contained 32% α-helix, 17% β-sheet, 20% β-turn, and 31% random coil. Based on this results, it has been concluded that the high percentage of random coil conformation estimated by this analysis suggests the presence of unstructured portions of the molecule alongside one or more folded regions (684). This conclusion was further supported by the 2D 1H NMR spectra (HSQC), which were shown to contain two main resonance types: well dispersed resonances typical of a folded conformation and sharp highly overlapped peaks typical of a random coil
conformation. Furthermore, limited proteolysis revealed that the intact protein was almost completely digested after 1 min of incubation with a series of proteases and a protease-resistant N-terminal domain was generated (684). These data indicated that ataxin-3 is composed of a structured N-terminal domain, followed by a flexible tail.

5.6.7. P/Q-type calcium channel α1A Subunit (CACNA1A)

The underlying mutation in SCA6, a dominantly inherited neurodegenerative disease characterized by progressive ataxia and dysarthria caused by cerebellar atrophy, is an expansion of the trinucleotide CAG repeat in exon 47 of the CACNA1A gene which encodes the α1A subunit of the P/Q type voltage-dependent calcium channel (685). Unlike many other polyglutamine diseases the expanded SCA6 alleles unusually have small expansions (21-30 repeats compared to generally >40 repeats in other polyglutamine diseases) (685). The product of the CACNA1A gene, P/Q-type Calcium Channel α1A Subunit (CACNA1A), is a protein with 2505 residues and a calculated molecular mass of 282.4 kDa. It has been found that the CACNA1A is processed in such a way that a C-terminal polyglutamine disease (686). Normal SCA7 alleles contain 4–35 CAG repeats, whereas pathological alleles expanded SCA6 alleles unusually have small expansions (21-30 repeats compared to generally >40 repeats in other polyglutamine diseases) (685). Unlike many other polyglutamine diseases the expanded SCA6 alleles unusually have small expansions (21-30 repeats compared to generally >40 repeats in other polyglutamine diseases) (685). The product of the CACNA1A gene, P/Q-type Calcium Channel α1A Subunit (CACNA1A), is a protein with 2505 residues and a calculated molecular mass of 282.4 kDa. It has been found that the CACNA1A is processed in such a way that a C-terminal polyglutamine-containing fragment which is less soluble and more toxic than the truncated polyglutamine stretch itself is produced (686). In one set of transcript variants, the (CAG)n-repeats occur in the 3' UTR, and are not associated with any disease. But in another set of variants, an insertion extends the coding region to include the (CAG)n-repeats which encode a polyglutamine tract. Expansion of the (CAG)n-repeats from the normal 4-16 to 21-28 in the coding region is associated with spinocerebellar ataxia 6 (687). This protein was predicted to have several long IDRs.

5.6.8. Ataxin-7

Spinocerebellar ataxia type 7 (SCA7) is characterized by cone-rod dystrophy retinal degeneration and is caused by a polyglutamine expansion within ataxin-7. It has been recently reported that ataxin-7 is a component of the mammalian STAGA (SPT3-TAF9-ADA-GCN5 acetyltransferase) transcription coactivator complex (688). In this context, ataxin-7 interacts directly with the GCN5 histone acetyltransferase component of STAGA, and mediates a direct interaction of STAGA with the CRX (cone-rod homeobox) transactivator of photoreceptor genes. Furthermore, poly(Q)-expanded ataxin-7 was incorporated into STAGA and inhibited the nucleosomal histone acetylation function of STAGA GCN5. Based on these results it has been suggested that the normal function of ataxin-7 may intersect with its pathogenic mechanism (688). Normal SCA7 alleles contain 4–35 CAG repeats, whereas pathological alleles contain from 36 to 306 CAG repeats (689). Ataxin-7 has 892 amino acids and a molecular mass of 95.4 kDa. However, at the SDS-PAGE this protein migrates at about 110 kDa (688). In other words, the apparent molecular mass of ataxin-7 determined by gel electrophoresis was found to be about 1.15-folds higher than that expected from amino acid sequence. This suggests that ataxin-7 possesses significant amount of intrinsic disorder.

5.6.9. TATA-Box-binding protein

SCA17 is characterized by the heterogeneous clinical phenotype, including ataxia, dementia, psychiatric symptoms, and, in some cases, epilepsy. Neurodegeneration in SCA17 is frequently widespread (atrophy of the striatum, thalamus, cerebral cortex, inferior olive, and nucleus accumbens have been reported), being most prominent in the cerebellum (690). Ubiquinated intranuclear inclusions were found in postmortem brain tissue from SCA17 patients as a result of immunohistochemical examination (690). SCA17 originates from the polyQ expansion of the TBP, which normally contains the polyQ tract of 25–42 glutamine residues, but is expanded >42 glutamines in SCA17 (690). TBP is required for transcriptional initiation by the three major RNA polymerases (RNAP I, II, and III) in eukaryotic nuclei. Being a component of distinct multi-subunit transcriptional complexes, TBP is involved in the expression of most eukaryotic genes (691). TBP is a 339 amino acids-long protein, which can be divided on two functional domains. The C-terminal domain is highly conserved among eukaryotes and mediates virtually all of the transcriptionally relevant interactions involving TBP (692), whereas the N-terminal domain is evolutionarily divergent and shows sequence conservation only in vertebrates. It has been demonstrated that polyQ expansion caused abnormal interaction of TBP with the general transcription factor TFIIIB and induced neurodegeneration in transgenic SCA17 mice (693). Furthermore, polyQ expansion was shown to reduce the in vitro binding of TBP to DNA. The mutated TBP fragments lacking an intact C-terminal DNA-binding domain were shown to be present in transgenic SCA17 mouse brains. PolyQ-expanded TBP with a deletion spanning part of the DNA-binding domain did not bind DNA in vitro but formed nuclear aggregates and inhibited TATA-dependent transcription activity in cultured cells (694). SDS-PAGE analysis of the murine TBP revealed that this protein is characterized by the apparent molecular mass of ~37 kDa, which exceeds the predicted molecular mass of 34.7 kDa (694). The difference between observed and calculated molecular masses was even higher for a truncated TBP fragment that lacks an intact C-terminal domain (694). Similarly, human TBP, a protein with the calculated molecular mass of 37.7 kDa, was shown to possess an apparent molecular mass of ~49 kDa (695).

5.7. ABri peptide and familial British dementia

The ABri is a 34 residue peptide that is the major component of amyloid deposits in familial British dementia (FBD), which is an autosomal dominant disorder with onset at around the fifth decade of life and full penetrance by age 60 characterized by the presence of amyloid deposits in cerebral blood vessels and brain parenchyma that coexist with NFTs in limbic areas (696). FBD patients develop progressive dementia, spasticity, and cerebellar ataxia. The protein subunit (termed ABri) is an example of an amyloid molecule created de novo by the abolishment of the stop codon in its precursor, a 266-amino acids-long type 2 transmembrane protein of unknown function (BRI-266) that is encoded by a single gene, BRI2, located on the long arm of chromosome 13 (697, 698). The FBD has a single
Protein intrinsic disorder in degenerative diseases

nucleotide change (TGA→AGA, codon 267) that results in an arginine residue substitution for the stop codon in the wild-type precursor molecule and a longer open reading frame of 277 amino acids in a disease-related protein (BRI-277 instead of BRI-266). The ABrI amyloid peptide is formed by the 34 C-terminal amino acids of the mutated precursor protein BRI-277, presumably generated from furin-like processing (699). Thus, the point mutation at the stop codon of BRI results in the generation of the 34 residue ABrI peptide (instead of the shorter 23 residue wild type peptide), which is deposited as amyloid fibrils causing neuronal dysfunction and dementia (700). It has been emphasized that although FBD and AD share almost identical neurofibrillary pathology and neuronal loss that co-localize with amyloid deposits, the primary sequences of the amyloid proteins (ABrI and Aβ) differ. Therefore, ABrI and Aβ amyloid deposition in the brain can trigger similar neuropathological changes (neuronal loss and dementia) and thus may be a key event in the initiation of neurodegeneration (700).

Using far-UV CD and NMR spectroscopy it has been recently established that ABrI is in the random coil-like conformation at slightly acidic pH (700). The solution pH was shown to play an important role in promoting the amyloid-like β-sheet structure and the characteristic fibril morphology of ABrI and this protein forms amyloid fibrils at pH 4.9 with no distinct fibril morphology being observed at neutral and slightly basic pH (pH 7.1–8.3), except for smaller spherical aggregates that gradually disappeared and assembled into larger amorphous aggregates (700). It has been also pointed out that at pH 4.9 the ABrI undergoes relatively slow β-aggregation, where it is possible for fibril formation to occur, similar to the behavior of the amyloid Aβ peptide (700).

5.8. ADan in familial Danish dementia

Familial Danish dementia (FDD) is a neurodegenerative disorder linked to a genetic defect in the BRI2 gene. Similar to FBD, FDD results from the genetic alterations in this gene and the deposited amyloid protein, ADan, is the C-terminal proteolytic fragment of a genetically altered BRI2 precursor molecule (701). The amyloid peptides ABrI and ADan originate as a result of two different genetic defects at, or immediately before, the BRI2 stop codon with a common final outcome in both diseases: regardless of the nucleotide changes, the ordinarily occurring stop codon is either non-existent (in FBD) or out of frame (in FDD) causing the genesis of an extended precursor featuring a C-terminal piece that does not exist in normal conditions (reviewed in (702)). ABrI and ADan are released by a furin-like proteolytic processing. Both these peptides are 34-residues-long, which share 100% identity on the first 22 residues, a completely different 12 amino acid C-terminus and have no sequence identity to any other known amyloid protein. Despite the structural differences among the corresponding amyloid subunits FDD and FBD show striking clinical and neuropathological similarities with AD, including the presence of NFTs, parenchymal amyloid and pre-amyloid deposits and CAA co-localizing with inflammatory markers, reactive microglia and activation products of the complement system (reviewed in (702)). Structural analysis revealed that similar to Aβ and ABrI, ADan is a typical natively unfolded protein, which is characterized by a random coil structure in a wide pH range and is prone to form fibrils in a pH-dependent manner (703).

5.9. Glial fibrillary acidic protein and Alexander disease

Alexander disease is a specific astrocytic disease caused by a dominant heterozygous mutation in glial fibrillary acidic protein (GFAP) (704, 705). A major pathological hallmark of Alexander disease is the presence of specific inclusion bodies called Rosenthal fibers (RFs) in astrocytes that are formed by the mutated GFAP (706). Besides mutated GFAP, these inclusions contains small heat shock proteins, including αB-crystallin and HSP27(707). Clinically, the phenotype of Alexander disease depends on the age of onset. The infantile form severely affects the entire central nervous system, with rapid progression and is characterized by megalencephaly, epilepsy, motor impairment, cognitive decline, and extensive loss of white matter with frontal predominance. However, the adult form progresses slowly and is characterized by predominant rhombencephalic degeneration without epilepsy, cognitive impairment, and little, if any, leukodystrophy. The juvenile form is intermediate in severity (708, 709). It has been shown that GFAP is characterized by an extremely high susceptibility to proteolysis (710, 711). Electrophoretic analysis of GFAP produced an apparent molecular mass of 54 kDa, which exceeds the calculated molecular mass of 49.9 kDa (710, 711). This aberrant electrophoretic mobility suggests that GFAP contains regions on intrinsic disorder.

5.10. Mitochondrial DNA polymerase γ and Alpers disease

Alpers disease, also known as progressive neuronal degeneration of childhood, is characterized by developmental regression, intractable epilepsy, progressive neurological deterioration, liver disease, and death usually before 10 years of age (712-714). Neuropathological changes include patchy neuronal loss and gliosis, particularly in the striate cortex (715), whereas the liver shows steatosis, cellular necrosis, focal inflammation, and fibrosis (716). Alpers disease is attributed to mutations in the catalytic subunit of the mitochondrial DNA (mtDNA) polymerase encoded by the polymerase γ gene (POLG1) (717). POLG is the only known DNA polymerase in the mitochondrion, which is responsible for ~1% of the total cellular DNA polymerase activity. The human POLG holoenzyme comprises a 140 kDa catalytic subunit (POLGα) and a 55 kDa accessory subunit (POLGβ). POLGα is a member of a DNA polymerase family with separate polymerase and 3’-5’ exonuclease domains thus exhibiting both DNA polymerase and 3’-5’ exonuclease activities. POLGβ increases DNA-binding affinity, stimulates the catalytic activities and enhances the processivity of the holoenzyme (718). The region of POLGα (444-820 fragment) that lies between the exonuclease and polymerase is known as spacer. Its size and sequence in POLGα are substantially different from those of other members of the DNA polymerase family. In POLGα, this large interdomain region is likely to
participate in DNA-template binding and guidance, as well as in subunit interactions. Importantly, spacer mutations were found frequently in the infantile Alpers syndrome, affecting most severely the brain and the liver (717, 719). These reports emphasize the exceptional variability of POLGα-associated neurological phenotypes and the specific role for spacer mutations in the most severe neurological manifestations (720). POLGα was shown to possess an apparent molecular mass of 145–147 kDa (based on the aberrant mobility in SDS-gel electrophoresis (721)), whereas its theoretical molecular mass is 139.5 kDa suggesting that POLGα contains a number of disordered regions.

5.11. DNA excision repair protein ERCC-6 and Cockayne syndrome

Cockayne syndrome (CS) (also known as Weber-Cockayne syndrome, or Neill-Dingwall Syndrome) is a rare, autosomal recessive disorder. Affected individuals suffer from postnatal growth failure resulting in cachectic dwarfism, photosensitivity, skeletal abnormalities, mental retardation and progressive neurological degeneration, retinopathy, cataracts and sensorineural hearing loss (722-724). Two complementation groups of CS (CS-A and CS-B) have been identified, the corresponding genes, CSA and CSB, have been cloned (725, 726) and their products biochemically characterized. The majority of CS cases are caused by defects in the CS complementation group B protein. CSA is a 44 kDa protein and belongs to the ‘WD repeat’ family of proteins (726), which exhibit structural and regulatory roles but no enzymatic activity. The CSB gene product is a 168 kDa protein (725), also known as DNA excision repair protein ERCC-6, belongs to the SWI/SNF family of proteins, which all contain seven sequence motifs conserved between two superfamilies of DNA and RNA helicases and which have roles in transcription regulation, chromosome stability, and DNA repair. The involvement of CSB in transcription, transcription-coupled repair of DNA, and base-excision repair might be simultaneous. However, it is suggested that some interregulation, depending on cellular status, takes place. This regulation is done via posttranslational modifications of CSB and changes in function and localization of its interaction partners. These many roles of CSB explain the multisystem manifestations of the CS phenotype (724).

In vivo studies demonstrate that CSB exists in a quaternary complex composed of RNA pol II, CSB, DNA and the RNA transcript. The CSB protein contains an acidic amino acid stretch (~60% of the residues in a 39-amino-acid stretch are acidic), a glycine-rich region and two putative nuclear localization signal (NLS) sequences (727). The cellular and molecular phenotypes of CS include increased sensitivity to oxidative and UV-induced DNA lesions. The CSB protein plays a crucial role in transcription-coupled repair. The corresponding CS-B cells are defective in the repair of the transcribed strand of active genes, both after exposure to UV and in the presence of oxidative DNA lesions (727). According to SDS-PAGE analysis, the CSB protein has an apparent molecular mass of ~200 kDa (727), whereas its theoretical molecular mass calculated from amino acid sequence is 168.4 kDa. This aberrant electrophoretic mobility suggests that CSB contains significant intrinsic disorder.

5.12. Survival of motor neurons protein and spinal muscular atrophy

Spinal muscular atrophy (SMA) is an autosomal recessive disease with a carrier frequency of about 1 in 50. SMA is the most common genetic cause of childhood mortality and leads to muscle weakness and atrophy due to the degeneration of motor neurons from the spinal cord (728). The gene responsible for disease is mapped to the survival of motor neurons (SMN) 1 (smn1) gene, which carries mutations in over 98% of all SMA patients (729). SMN variants (SMNΔelta7 and SMN-Y272C) found in patients with SMA not only lack antiapoptotic activity but also are potent proapoptotic, causing increased neuronal apoptosis and animal mortality. The SMN protein is a part of a larger protein complex that is present both in the nucleus and the cytoplasm. In the nucleus, SMN protein localizes to spots that are rich in small nuclear ribonucleoprotein particles (snRNPs). In the cytoplasm, the SMN protein plays an important role in the assembly of these snRNPs (730). The SMN protein interacts with core components of the snRNPs, Sm proteins. SMA-causing mutations in a C-terminal region and in the central Tudor domain of the SMN protein have been shown to affect the Sm interaction. Mutations in the C-terminal region may interfere with the Sm interaction indirectly, since this region is also required for SMN protein oligomerization (731). The SMN protein Tudor domain has been shown to directly bind to the arginine-glycine (RG) rich tails of the Sm proteins in vitro (732, 733). Furthermore, the type I SMA causing point mutation E134K in the SMN protein Tudor domain abolishes Sm binding in vitro and interferes with snRNPs assembly in vivo (733).

In vivo the RG-rich tails of the SmB, SmD1 and SmD3 proteins are post-translationally modified and contain symmetrically dimethylated arginine residues (734, 735). This modification strongly enhances the affinity of the SMN/Sm interaction and has been implicated in the regulation of uridine-rich snRNP assembly (735, 736). Many other proteins, including coilin, RNA helicase A, fibrillarin and heterogeneous nuclear ribonucleoproteins, interact with the SMN complex and contain RG-rich domains that can potentially be methylated (737, 738), suggesting that the SMN protein Tudor domain could have an additional function in the regulation of these interactions. The crystal structure of the SMN protein Tudor domain comprising residues 82-147 was solved to high (1.8 Å) resolution (739, 740). The crystal structure consists of a five-stranded β-sheet that forms a β-barrel. Comparison of the crystal structure and an NMR structure revealed that the backbone conformation of both structures is very similar. However, differences were observed for the cluster of conserved aromatic side-chains in the symmetrically dimethylated arginine residues (sDMA) binding pocket, suggesting that the SMN protein Tudor domain adopts two different conformations in the sDMA binding pocket (740). Using the SDS-PAGE analysis, full-length SMN protein (calculated molecular mass 31.8 kDa)
was shown to possess an apparent molecular mass of 39 kDa (741), suggesting that a noticeable fraction of this protein is intrinsically disordered.

5.13. Septins in neuropathology

Septins are ubiquitous, evolutionarily conserved GTP-binding proteins that can form homo- and heteromeric complexes to assemble into filaments that may serve as structural scaffolds for the assembly of other proteins (742, 743). Septins are the members of the family of cell division cycle (Cdc) genes. They are evolutionary conserved across the eukaryotic phylogeny from unicellular yeast to complex metazoa including Homo sapiens (744), and constitute a group of GTP-binding and filament-forming proteins that belong to the large superclass of P-loop GTPases (745). In humans, there are at least 12 septin genes encoding septins SEPT1-12. Most of them undergo complex alternative splicing with some degree of tissue specificity (745). Analysis of septin amino acid sequences revealed potential motifs and regions, some of which are conserved among the currently known orthologues. For example, the septin family of genes possesses a conserved GTP-binding domain, and they fall into the large superclass of P-loop GTPases (746). This central GTP-binding domain is highly conserved in all human septins, with 58% identity or similarity. All septins have a P-loop (747, 748) that is defined by the Walker A motif (GxxGxGKST), Walker B motif (DxxG), and the GTP-specificity motif (xKxD). Septins lie in the TRAFAC subclass of P-loop GTPases defined by a conserved threonine required for hydrolysis of the triphosphate moiety (746). Contrarily to this highly conserved GTP-binding domain, the N- and C-terminal domains vary greatly both within and between species. In fact, the N-terminal domain of septins contains a polybasic region, which is somewhat conserved through eukaryotic phylogeny. The C-termini of the known human septins are more diverse. The diversity in the N- and C-terminal regions flanking the core of the 12 known human septins is combined with the extensive alternate splicing allowing septin genes to encode multiple isoforms (745).

Originally, septins were identified based on the cell division cycle mutants with defects in cytokinesis (749). Now, it is established that septins are heavily involved in various biological processes inside the cell. They were shown to have diverse biological roles including cytokinesis (750-752), cell polarity determination and maintenance (753-755), cell movement and membrane associations (756), vesicle trafficking (757), exocytosis (758, 759), and apoptosis (760). Septins can interact with both microtubules and actin (750, 761-763), potentially playing a role as adaptors between the two cytoskeletons and as regulators of processes in which both actin and microtubules are involved (764). Septins are proposed to be involved in several microtubule-dependent processes, including karyokinesis, exocytosis, and maintenance of cell shape (764). Various septins have been shown to colocalize or interact with the microtubule cytoskeleton. Therefore they might play an important role in regulation of the microtubule dynamics via specific interaction with microtubule-associated proteins modulating microtubule stability (764). Recently, the structural properties of the SEPT4 were analyzed by a number of biophysical techniques, including native gel electrophoresis, CD spectroscopy, fluorescence spectroscopy, DLS, and SAXS as well as with bioinformatics tools (765). To this end, the full-length form of human SEPT4 and its individual N-terminal, GTPase, and C-terminal domains were expressed in E. coli and purified. Biophysical analysis revealed that the N-terminal domain behaves as a typical IDP, containing little regular secondary structure. The central GTPase domain was catalytically active and represented a mixed α/β structure, probably based on an open β-sheet. The C-terminal domain was shown to form homodimers and can be divided into two regions, the second of which is α-helical and consistent with a coiled-coil structure (765).

Septins are involved in several neurological disorders. This conclusion follows from several observations including the brain-specific expression of some septins, the differential regulation of septins in neural development, and the association of septins with some disease states or pathological hallmarks (745). For example, septins were found in neurofibrillary tangles in Alzheimer's disease (766). In fact, SEPT1, 2, and 4 were shown to be associated with tau-based helical filaments. It has been even proposed, that septins contribute to the formation of tangles and therefore are pathogenetically significant. Although several human septins are expressed exclusively or predominantly in the nervous system, their expression is under the strict spatial (767) and temporal (768) regulation. Furthermore, since the septin expression was linked with exocytosis (758, 759) and since septins were shown to form complexes with syntaxin and other secretion-associated molecules, it has been proposed that septins can be associated with the critical function of secretion via the exocyst complex and can be involved in vesicle trafficking (745). SEPT2 is included into the exocyst complex (757) and modification of its GTP binding activity was shown to be accompanied by the altered neurite sprouting (769). This septin has been also reported to be overexpressed in brain tumors (770).

Based on the results of the proteomic analysis the abnormalities in the septin expression were found in Down's syndrome (771). SEPT5, a v2 has been reported to be a parkin-binding protein and parkin can function as an E2-dependent ubiquitin ligase capable of promoting the degradation of SEPT5 (772). SEPT5 was shown to accumulate in the brains of individuals with autosomal recessive juvenile Parkinsonism (773), whereas SEPT4 has been found in neurofibrillary tangles of Alzheimer disease brains and in α-synuclein-positive cytoplasmic inclusions in Parkinson disease brains (774).

5.14. Neurotrophin NGF and neurodegeneration

The neurotrophin family of growth factors consists of nerve growth factor (NGF), brain derived neurotrophic factor (BDNF), and neurotrophins-3 and -4 (NT-3, NT-4) and regulates neuronal survival and synaptic plasticity (775). Neurotrophins are synthesized as precursors (proneurotrophins) that are proteolytically cleaved to mature, biologically active neurotrophins (776).
Figure 8. \textit{Ab initio} models for the shape determination of proNGF. The models obtained by DAMMIN are shown by grey spheres. These models are superimposed on the Cα traces obtained with BUNCH for the pro-peptide region (light grey); in black, the Cα traces of the crystallographic structure of mouse NGF (PDB code 1BET) are shown. (A) Crab-like shape; (B) rod-like shape. Reproduced with permission from (783).

For example, the neurotrophin NGF is translated as a pre-pro-protein [pre-proNGF (precursor of NGF with signal peptide)] of 27 kDa, containing a signal peptide for protein secretion (pre-peptide) and the precursor protein (proNGF, NGF precursor, without signal peptide). ProNGF is cleaved by the convertase furin to produce the mature NGF. Mature neurotrophins selectively bind to members of the Trk family of receptor tyrosine kinases, but they also interact with a structurally distinct receptor, pan-neurotrophin receptor p75NTR. Although p75NTR can increase the affinity and specificity of Trk-neurotrophin interactions, p75NTR can also induce apoptosis in oligodendrocytes, neurons, and vascular smooth muscle cells when Trk activation is reduced or absent (777, 778).

NGF is involved in the maintenance and growth of neurons (779). ProNGF was found to be a high-affinity ligand for p75NTR and was shown to induce p75NTR-dependent apoptosis (777, 778). The specific receptor for the proNGF is sortilin (780). ProNGF has also been found to bind to the high-affinity NGF receptor TrkA (tropomyosin receptor kinase A) and to induce the survival-signaling pathway, although it is less efficient than the mature NGF (781). ProNGF was shown to be the predominant form of NGF in mouse, rat, and human brain tissue, whereas little or no mature NGF was detected (782).

Interestingly, a high-throughput crystallization screen of proNGF failed to produce crystals, suggesting that the high flexibility of the pro-part of proNGF might influence its crystallization propensity (783). In agreement with this hypothesis, solution small angle X-ray scattering measurements revealed that proNGF was dimeric and appears to have two equally populated structures (see Figure 8), a globular crab-like form and an elongated rod-like form, pointing to an intrinsically disordered pro-region of NGF (783). It has been also emphasized that these two models provide grounds for the interpretation of the available biological data for proNGF.

In AD, neuronal dysfunction and degeneration occur in the basal forebrain cholinergic neurons (BFCN), with reduction in neocortical cholineacetyltransferase (ChAT) activity (784-786). This reduced ChAT activity correlates with the degree of dementia and thus has been regarded as a principal factor associated with the memory
loss characteristic of AD (787). NGF, which is synthesized within BFCN target regions, such as the hippocampus and cerebral cortex (788), maintains survival of the BFCN after injury and regulates cholinergic neurotransmitter levels (789-792). Blocking NGF availability to BFCN results in memory deficits (793, 794). This led to the theory that BFCN degeneration in AD is the result of a deficit in NGF (795). Decreased NGF immunoreactivity in the basal forebrain of AD patients and increased NGF protein in the cerebral cortex and hippocampus have been demonstrated using bioassay and ELISA (796-800). These changes and the decreased trkA expression in AD (801-805) were consistent with a defect in the NGF transport in AD brain. Alternatively, the availability of NGF protein to BFCN can be reduced by defects in posttranslational modification of NGF. This, in turn, might result in the decrease in the ChAT and trkA levels. In agreement with this hypothesis, a twofold increase in proNGF in AD parietal cortex compared to controls was found, indicating that this precursor form preferentially accumulates in AD (782).

5.15. FUS and TDP-43 in amyotrophic lateral sclerosis and frontotemporal lobar degeneration

Amyotrophic lateral sclerosis (ALS, also known as Lou Gehrig’s disease) is a fatal human neurodegenerative disease affecting primarily motor neurons, selective degeneration of some of these neurons results in gradual muscle weakness and atrophy that ultimately leads to death (806), where the degeneration of motor neurons is linked to aggregation of two RNA-binding proteins, TDP-43 (TAR DNA-binding protein of 43 kDa) and FUS (fused in sarcoma) (807), that are abnormally deposited in neuronal and glial cytoplasmic inclusions. These two proteins were also found to label pathological cytoplasmic inclusions in some patients with frontotemporal lobar degeneration (FTLD) (808), which is the second most common dementia below the age of 65 years and is characterized by atrophy of the frontal and temporal lobe (809). More than 40 dominant mutations in the TARDBP gene encoding TDP-43 have been identified in ALS and in some FTLD patients (807, 810). Furthermore, some FUS mutations are also associated with familial forms of ALS and FTLD (811-813).

The biological roles of TDP-43 are mostly RNA-centered, and this protein has several thousand RNA targets in the brain (814, 815), is involved in microRNA processing (816, 817) and pre-mRNA processing (818). TDP-43 has two RNA recognition motifs (RRMs) (818), and a C-terminal glycine-rich domain that mediates protein–protein interactions (819-821), being intrinsically disordered (808) and showing similarity to yeast prions (55, 808, 823). Importantly, almost all disease-associated mutations in this protein are clustered within this intrinsically disordered C-terminal glycine-rich domain (808).

FUS is also a DNA/RNA-binding protein that regulates transcription and splicing of hundreds of target genes (824-826). FUS contains several RNA-binding elements, such as arginine-glycine-glycine (RGG) domains, an RRM and a zinc finger (808), and possesses an N-terminal serine-tyrosine-glycine-glutamine (SYGQ)-rich transcriptional activation domain (827, 828), which is intrinsically disordered and is predicted to have prion-like properties (55, 808, 823).

5.16. HIV-associated dementia, aggregation of Fas and upregulation of Bad and Bax

One of the consequences of the human immunodeficiency virus-1 (HIV-1) infection is a neurological disease culminating in HIV-associated dementia or HAD seeing in 20-40% of patients infected with HIV-1 (829). It is believed that HAD is a result of the neuronal apoptosis induced by the HIV viral coat glycoprotein gp120 (830-832) via initiation of the gp120-triggered apoptotic mitochondrial membrane permeability (833). Importantly, similar to many other neurodegenerative diseases, pathogenesis of HAD is associated with mitochondrial dysfunction and oxidative damage (833). One of the pathways in this HIV-induced apoptosis is associated with the death receptor, Fas (834, 835), aggregation of which leads to the activation of Fas-associated protein with death domain (FADD) and caspase-8, and results in the formation of death-induced signaling complex (DISC) (833) that includes the death receptor FAS (mean PONDR FIT score of 0.140), death ligand (mean PONDR FIT score of 0.448) and adaptor proteins such as FADD (mean PONDR FIT score of 0.274) and TRADD (mean PONDR FIT score of 0.298). Fas is a member of the TNF receptor family, containing 3 cysteine-rich domains (CRDs) in its extracellular regions, a single-path transmembrane domain, and a C-terminally located death domain which is positioned within the intracytoplasmic region and is separated from a transmembrane domain by a long disordered loop. In addition to the activation of the Fas pathway, treatment of the SH-SY5Y neuroblastoma cells with gp120 showed an increase in expression of the important proapoptotic proteins Bad and Bax (mean PONDR FIT score of 0.245) were shown to be upregulated (833). It was shown that similar to other BH3-only proteins, which are members of the BCL-2 family that serve as key initiators of programmed cell death BAD is largely disordered in solution (836).

5.17. Valosin-containing protein in neurodegeneration

Mutations in valosin-containing protein (VCP, also known as p97, transitional endoplasmic reticulum ATPase, or a major cytosolic AAA (ATPase associated with a variety of cellular activities)) are known to be associated with the development of inclusion body myopathy with early-onset Paget disease and frontotemporal dementia (IBMPFD) and amyotrophic lateral sclerosis 14, with or without frontotemporal dementia (ALS14) (837-842). Clinically, IBMPFD (which is an autosomal dominant disorder) resembles limb girdle muscular dystrophy being characterized by adult-onset proximal and distal muscle weakness. Additionally, in some cases IBMPFD is connected with early-onset Paget disease of bone and premature frontotemporal dementia (843, 844). Therefore, IBMPFD is characterized by incomplete penetrance of three main pathological features: disabling muscle weakness (in 90%), osteolytic bone lesions consistent with Paget disease (in 51%), and
Protein intrinsic disorder in degenerative diseases

frontotemporal dementia (in 32%) (844). ALS14 is an inherited neurodegenerative disorder caused by VCP mutations that affects upper motor neurons in the brain and lower motor neurons in the brain stem and spinal cord, resulting in fatal paralysis. Some patients with ALS14 are known to develop frontotemporal dementia (842). Therefore, both IBMpFD and ALS14 are caused by mutations in VCP, and in some families with a VCP mutation, some family members may have ALS14, and other members may have IBMpFD. At the molecular level, it was shown recently that the VCP K524A mutant and the triple VCP mutant R93C-R155C-K524A were able to trigger Huntingtin aggregate formation, block protein degradation via the proteasomal pathway, and make cells highly susceptible to ER stress-induced cell death (845).

VCP contains the N-terminal domain (N-domain), the adjacent AAA domain (D1) and the intrinsically disordered C-terminal domain. Structural analysis on the VCP N-D1 fragment bearing IBMpFD-related mutation at the interface between the N and D1 domains revealed that the transition from the ADP- to the ATP-S-bound state is accompanied by a loop-to-helix conversion in the N-D1 linker and by an apparent reordering in the N-terminal region of p97 (841). Although the C-terminal domain lacks structural description, it is predicted to contain long disordered region and is known to possess multiple PTM sites.

5.18. Acetyltransferase p300 in Parkinson’s disease

Acetyltransferase p300 plays an important role in the execution of multiple biological programs, such as differentiation, senescence, apoptosis, and is involved in transforming activity of several viruses, with the vast majority of these functions taking place in the nucleus, via regulation of transcription and promotion of acetylation of many proteins (846, 847). Among acetylation targets of p300 are all four histones in nucleosome, SIRT2, ALX1, many proteins (846, 847). Among other functions ascribed to this protein, optineurin regulates endocytic trafficking of transferrin and its receptor to the recycling endosomes and is involved in regulation of the transcription factor NF-κB (852). C-terminal half of optineurin shows considerable homology with the NF-κB essential modulator (NEMO) and contains ubiquitin-binding domain and zinc finger domain (852). Although structural information is not available for the entire protein, the structure of the C-terminal zinc finger domain is known (residues 550-577, PDB ID: 2L04, see Figure 9A). Bioinformatics analysis revealed that this protein is expected to be mostly disordered (see Figure 9B).

6. SECOND D³: DYSFUNCTION OF NEURODEGENERATION-PROMOTING IDPS

Dysfunction and dysregulation of IDPs related to the pathogenesis of various neurodegenerative diseases can be a subject of the multivolume book. Since some of the aspects of these dysfunction and dysregulation were covered in sections dealing with individual neurodegeneration-related proteins (see above), there is no need to reproduce them here. However, one should keep in mind that neurodegenerative diseases are a subset of so-called conformational maladies since they are characterized by the conformational changes, misfolding, and often aggregation of an underlying protein. To better understand these conformational diseases, perturbed protein functionality have to be considered, since the high level of intrinsic disorder in pathogenic proteins is a clear reflection of the involvement of these proteins in some crucial signaling
functions. Therefore, the dysregulation and misfolding of these proteins can result in their dysfunction, ultimately leading to the development of life-threatening pathological conditions. In other words, mutations and/or changes in the environment may result in protein confusion, reducing its capability to recognize proper binding partners, and leading to the formation of nonfunctional complexes and deadly aggregates (349).

7. THIRD D³: IDPS CONTROLLING AND REGULATING THE IDPS RELATED TO THE NEURODEGENERATION

An interesting twist in the D³ story is that neurodegeneration-related IDPs are not only abundantly found in neurodegenerative diseases and cause these disorders by dysfunction and dysregulation, but also that they are regulated and controlled by other IDPs. Although there are many examples of such disordered regulation of disordered proteins in neurodegenerative disorders, only four of them are considered below.

7.1. Disordered chaperones in neurodegeneration: Malleable guardians

Protein misfolding and aggregation represent two obvious outcomes of distorted proteostasis. A complex protective system is elaborated by Nature to overcome these distortions in proteostasis, with a network of molecular chaperones being the major component of this cellular quality control system (853-856). Chaperones help maintaining an intricate balance between protein synthesis and degradation and play a number of fundamental roles in protection of cells from the devastating consequences of uncontrolled protein misfolding and aggregation via several intricate mechanisms (225). These mechanisms can be grouped into three major classes of action: prevention,
reversal and elimination, and therefore chaperones help proteins to fold, prevent them from undergoing misfolding and aggregation, and work with misfolded and aggregated species to promote their disaggregation and unfolding followed by either productive refolding or degradation. In other words, chaperones are arranged into several functional subclasses based on their molecular mechanism of action, such as unfolded chaperones, which utilize ATP-dependent conformational changes to promote unfolding and subsequent refolding of their substrates; holding chaperones, which retain their partially folded/misfolded substrates and prevent them from subsequent aggregation; and disaggregating chaperones, which solubilize the aggregated proteins. At the elimination step, chaperones target misfolded proteins for degradation by the ubiquitin-proteasome system and/or the autophagy-lysosome system (857). The increased levels of misfolding and aggregation result in the abuse and potential failure of the quality control system. In its turn, the failure of this protein quality control system to fulfill its functions or malfunction of its components generates the potential for tissue-specific build-up of protein aggregates, and is related to the development of neurodegenerative diseases (858). Chaperone network is vast and involves the highly coordinated and orchestrated action of multiple players (225).

The analysis of the abundance of predicted IDRs in chaperones revealed a high proportion of such regions in these proteins, with ~40% of their residues falling into disordered regions and ~15% falling within disordered regions longer than 30 consecutive residues (343). A recent comprehensive review analyzed the various roles of intrinsic disorder in function of many chaperones and clearly showed that many neuroprotective chaperones/co-chaperones are either completely disordered or possess long disordered regions (225). IDRs determine the promiscuity of chaperones, act as pliable molecular recognition elements, wrap misfolded chains, and participate in disaggregation and local unfolding of the aggregated and misfolded species. IDRs play important roles in precise orchestration of coordinated actions of chaperones, co-chaperones, and decorating proteins, which often operate as large chaperone machines, and which communicate with each other to form sophisticated chaperone networks (225). Based on these observations it has been concluded that protein intrinsic disorder plays a crucial role in the coordination and regulation of these chaperone machines and networks, thus helping form a flexible net of malleable guardians (225).

7.2. Sirtuins in neurodegenerative diseases

Sirtuins constitute an important family of regulatory proteins involved in several physiological functions including control of gene expression, metabolism, and aging (859). In mammals, there are seven sirtuins (Sirt1 to Sirt7), with the most studied being the SIRT1 (860). Sirtuins act as deacetylases and mono-ADP-ribosil-transferases, belonging to class III histone deacetylases that differs from the other classes by their mechanism of action being dependent on NAD⁺ (861, 862), that have a number of important roles in the body’s response to various kinds of stress and toxicity (859) and that are able to affect biological aspects involved in metabolic diseases and aging in mammals (861). Furthermore, members of this protein family are highly expressed in various regions of the brain involved in cognitive functions and have been demonstrated to regulate cellular protection against oxidative stress in many diseases that involve aging and neurodegeneration (863). Therefore, modulation of sirtuins represents an attractive direction for potential mitigation of the effects and complications of aging and neurodegenerative diseases (861).

In PD, the NAD-dependent deacetylase Sirt1 is involved in the regulation of autophagy (864) that is responsible for the clearance of aggregated α-synuclein (859). Overexpression of Sirt1 in animal models of PD reduced the α-synuclein aggregation (865) likely via that activation of molecular chaperones (866).

In AD models, such as in vitro cell cultures and transgenic mouse models, SIRT1 was shown to attenuate amyloidogenic processing of APP via the increase in the production and activity of an enzyme responsible for the non-amyloidogenic cleavage of APP, α-secretase, through activation of the α-secretase gene ADAM10 (867). Since the pathological accumulation of the amyloidogenic Aβ species results from the β-secretase and γ-secretase activities, the mentioned α-secretase upregulation shifts APP processing towards the non-amyloidogenic Aβ forms (867). Also, deactivation of Sirt1 was shown to result in increased levels of acetylated and pathogenic phosphorylated forms of tau protein, likely due to the blockage of the proteasome-mediated degradation (868).

In mouse models of HD, the HD-related brain pathology is enhanced in the brain-specific Sirt1 knockouts, whereas overexpression of Sirt1 improves survival, enhances the BDNF expression, and decreases neuropathology (869). This neuroprotective role of Sirt1 relies on one of the brain-specific modulators of CREB activity, the CREB-regulated transcription coactivator 1 (TORC1). In the norm, TORC1 is activated via the Sirt1-controlled deacetylation that promotes dephosphorylation of TORC1 and enhances its interaction with CREB (869). Although mutant huntingtin protein interferes with the TORC1–CREB interaction and represses the BDNF transcription, this defect is fixed in vitro and in vivo by Sirt1 overexpression (869).

Therefore, Sirt1 definitely acts as a master-regulator of various neuropathologies exacerbated by several neurodegeneration-promoting IDPs (α-synuclein, tau, Aβ, and huntingtin). Importantly, recent comprehensive analysis of the sirtuin family by a wide array of bioinformatics tools revealed that these proteins in addition to the highly similar catalytic cores are characterized by the presence of long disordered terminal arms (870). Furthermore, it has been concluded that these disordered tails of sirtuins have crucial biological roles since “the terminal segments of the majority of sirtuins possess a number of structural features and chemical and
Protein intrinsic disorder in degenerative diseases

physical properties that strongly support their involvement in activities of recognition and interaction with other protein molecules” (870).

7.3. Stress granules, regulated protein aggregation, and neurodegeneration

An interesting recent development is the recognition of the importance of stress granules (SGs) for the pathology of neurodegenerative diseases (808, 871). SGs are cytoplasmic foci that represent a special form of RNA granules containing several RNA-binding proteins and RNA molecules stalled at the pre-initiation stage and that are formed as a specific cellular response to the stressful conditions. In other words, SGs are the cytoplasmic messenger ribonucleoprotein (mRNP) particles that possess defined cytoprotective function (808).

SGs sequester and silence mRNAs encoding house-keeping proteins, modify local patterns of translation within the cell, sequester signaling molecules that regulate cell viability (872, 873), and shift RNA translation towards cytoprotective proteins, such as heat shock proteins and chaperones (874). Among regulatory proteins sequestered by SGs are some apoptotic regulators (e.g. TRAF2 and RACK1) (875-877) and a regulator of cell growth and metabolism the mammalian target of rapamycin complex 1 (mTORC1) (878, 879). In addition to mRNA, SGs contains mRNA-bound 48S pre-initiation complexes composed of small ribosomal subunits and translation initiation factors (e.g. eIF3, eIF4E and eIF4G). Furthermore, SGs contain proteins involved in mRNA stabilization, processing and transport, such as PABP-1, T cell internal antigen-1 (TIA-1), TIA-1-related (TIAR) and Ras-GTPase-activating protein SH3-domain-binding protein (G3BP) (808, 877, 880-884). Contrarily to amyloid or amyloid-like fibrillation, the process of GS formation is completely reversible and is tightly controlled (871). The formation of SGs is a very complex process, the complete description of which is definitely outside the scopes of this review. In brief, SG nucleation is initiated by several RNA-binding proteins, with the most commonly examined being TIA-1, TIAR, TTP, and G3BP (872), whereas SG maturation involves incorporation of many pro-apoptotic proteins (e.g., TRAF2, ROCK1, and RACK1) and other regulators of apoptosis (such as RSK2 and FAST kinase), leading to the inhibition of the apoptotic response (871). Among apoptosis-unrelated signaling and regulating proteins incorporated into the SGs are JNK, MKK7, rhoA, AKAP350A, WDR62, and HDAC6 (871).

Relation of SGs to neurodegeneration comes from the facts that:

(a) SGs are co-localized with insoluble protein aggregates in many neurodegenerative diseases (871);
(b) SGs frequently contain RNA-binding proteins related to the pathogenesis of various neurodegenerative diseases, such as TDP-43 and FUS, related to the pathology of ALS and FTLD, SMN related to the SMA pathology, SCA2-related ataxin-2,

7.4. Tubulin polymerization promoting protein and α-synuclein aggregation

The neuropathological hallmark of MSA (which is a progressive neurodegenerative disorder presenting variable combinations of Parkinsonism, cerebellar ataxia, corticospinal and autonomic dysfunction) is the α-synuclein-immunopositive glial cytoplasmic inclusions (GCIs). Pathological aggregation of α-synuclein in oligodendroglia is promoted by the tubulin polymerization promoting protein (TPPP)/p25, which specifically accelerates α-synuclein oligomer formation and co-immunoprecipitates with α-synuclein (888). The major function of TPPP is in the maintenance of the microtubule network integrity, where TPPP is involved in regulation and control of the polymerization of tubulin into microtubules, microtubule bundling and the stabilization of existing microtubules (889, 890). This protein may be also be involved in mitotic spindle assembly and nuclear envelope breakdown (889), and play a crucial role in the myelination of oligodendrocytes (891). TPPP is highly disordered, widely expressed, possesses multiple PTM sites, and is involved in multiple interactions with unrelated partners (see Figure 10). Structural analysis revealed that TPPP/p25 is a typical IDP that partially folds as a result of Zn⁺ binding forming a molten globule-like structure (891).

8. CONCLUSIONS

Intrinsic disorder is highly abundant among proteins associated with human neurodegenerative diseases. This provides a strong factual support to a D² (disorder in disorders) concept (892). The validity of this concept in neurodegeneration is illustrated at several levels, starting from the results of the bioinformatics analysis of an extended set of proteins associated with various neurodegenerative conditions and ending with the extensive data for a number of well-characterized neurodegeneration-related proteins. High degree of association between intrinsic disorder and neurodegenerative diseases is due to the unique structural and functional peculiarities of IDPs and IDR.s. IDPs/IDRs are among major cellular regulators, recognizers and signal transducers. Their functionality and misbehavior are modulated via a number of post translational modifications (i.e., tau protein). Many IDPs/IDRs can fold (completely or partially) upon interaction with corresponding binding partners. They
Figure 10. Tubulin polymerization promoting protein as an intrinsically disordered hub controlling α-synuclein aggregation. A. PONDR-FIT score distribution within the sequence of TPPP/25 (UniProt ID: O94811). B. Interactome of TPPP/p25 analyzed by the STRING database (accessible at http://string-db.org) and shown as the network of predicted associations for a particular group of proteins. The network nodes are proteins, whereas the edges represent the functional associations evaluated based on the experiments, search of databases and text mining. Different line colors of edges represent different types of evidence for the association, such as experimental/biochemical data (pink lines), association in curated databases (blue lines), and co-mentioned in PubMed abstracts (dark yellow lines).
Protein intrinsic disorder in degenerative diseases

possess multiple binding specificity and they are able to participate in one-to-many and many-to-one interactions. Disordered proteins in neurodegenerative diseases constitute a novel D³ paradigm, which has at least three levels: (a) IDPs are common in neurodegenerative diseases; (b) pathogenesis of these diseases is connected to dysfunction of corresponding IDPs; and (c) disease-related IDPs are controlled by other IDPs.

9. ACKNOWLEDGEMENTS

This work was supported in part the Program of the Russian Academy of Sciences for the "Molecular and cellular biology".

10. REFERENCES


24. J. S. Pedersen, D. Dikov, J. L. Flink, H. A. Hjuler, G. Christiansen and D. E. Otzen: The changing face of...
Protein intrinsic disorder in degenerative diseases


Protein intrinsic disorder in degenerative diseases


Protein intrinsic disorder in degenerative diseases


80. C. Blake and L. Serpell: Synchrotron X-ray studies suggest that the core of the transthyretin amyloid fibril is a continuous beta-sheet helix. Structure, 4(8), 989-98 (1996)


Protein intrinsic disorder in degenerative diseases


Protein intrinsic disorder in degenerative diseases


146. V. N. Uversky: Intrinsic disorder in proteins associated with neurodegenerative diseases. Front Biosci, 14, 5188-238 (2009)

147. A. Kurz and R. Perneczky: Novel insights for the treatment of Alzheimer's disease. Prog...
Protein intrinsic disorder in degenerative diseases

*Neuropsychopharmacol Biol Psychiatry*, 35(2), 373-9 (2011)


Protein intrinsic disorder in degenerative diseases

and rescuing the hypoacetylation levels of histone H3 and H4. *PLoS One*, 8(1), e54792 (2013)


192. S. Longhi and M. Oglesbee: Structural disorder within the measles virus nucleoprotein and phosphoprotein. *Protein Pept Lett*, 17(8), 961-78 (2010)

193. T. H. Reingewertz, D. E. Shalev and A. Friedler: Structural disorder in the HIV-1 Vif protein and
Protein intrinsic disorder in degenerative diseases

interaction-dependent gain of structure. *Protein Pept Lett*, 17(8), 988-98 (2010)


Protein intrinsic disorder in degenerative diseases


249. N. Srinivasan and S. Kumar: Ordered and disordered proteins as nanomaterial building blocks. Wiley
Protein intrinsic disorder in degenerative diseases

Interdisciplinary Reviews-Nanomedicine and Nanobiotechnology, 4(2), 204-218 (2012)


254. L. Breydo and V. N. Uversky: Role of metal ions in aggregation of intrinsically disordered proteins in neurodegenerative diseases. Metallomics, 3(11), 1163-80 (2011)


Protein intrinsic disorder in degenerative diseases


Protein intrinsic disorder in degenerative diseases


Protein intrinsic disorder in degenerative diseases


Protein intrinsic disorder in degenerative diseases


Protein intrinsic disorder in degenerative diseases


400. L. A. Munishkina, A. Ahmad, A. L. Fink and V. N. Uversky: Guiding protein aggregation with macromolecular crowding. Biochemistry, 47(34), 8993-9006 (2008)


406. Z. Serber and V. Dotsch: In-cell NMR spectroscopy. Biochemistry, 40(48), 14317-23 (2001)


Protein intrinsic disorder in degenerative diseases


Protein intrinsic disorder in degenerative diseases


Protein intrinsic disorder in degenerative diseases


Protein intrinsic disorder in degenerative diseases


Protein intrinsic disorder in degenerative diseases


241
Protein intrinsic disorder in degenerative diseases


Protein intrinsic disorder in degenerative diseases


Protein intrinsic disorder in degenerative diseases


Protein intrinsic disorder in degenerative diseases


Protein intrinsic disorder in degenerative diseases


Protein intrinsic disorder in degenerative diseases


668. S. Chen, V. Berthelier, W. Yang and R. Wetzel: Polyglutamine aggregation behavior *in vitro* supports a
Protein intrinsic disorder in degenerative diseases


Protein intrinsic disorder in degenerative diseases


Protein intrinsic disorder in degenerative diseases


723. E. C. Friedberg: Cockayne syndrome—a primary defect in DNA repair, transcription, both or neither? Bioessays, 18(9), 731-8 (1996)


736. W. J. Friesen, S. Massenet, S. Paushkin, A. Wyce and G. Dreyfuss: SMN, the product of the spinal muscular atrophy gene, binds preferentially to dimethylarginine-containing protein targets. Mol Cell, 7(5), 1111-7 (2001)


Protein intrinsic disorder in degenerative diseases


Protein intrinsic disorder in degenerative diseases


Protein intrinsic disorder in degenerative diseases


Protein intrinsic disorder in degenerative diseases


Protein intrinsic disorder in degenerative diseases


847. S. Tyteca, G. Legube and D. Trouche: To die or not to die: a HAT trick. Mol Cell, 24(6), 807-8 (2006)


**Protein intrinsic disorder in degenerative diseases**


887. Y. Furukawa, K. Kaneko, G. Matsumoto, M. Kurosawa and N. Nukina: Cross-seeding fibrillation of Q/N-rich proteins offers new pathomechanism of...
Protein intrinsic disorder in degenerative diseases


Abbreviations: AAA, ATPase associated with a variety of cellular activities;ACHé, acetylcholinesterase;AD, Alzheimer's disease;AFM, atomic force microscopy;ALS, amyotrophic lateral sclerosis;ANS, 8-anilino-1-naphthalenesulfonic acid; APP, amyloid ?-protein precursor; APR, aggregation-prone region;BDNF, brain derived neurotrophic factor;BSE, bovine spongiform encephalopathy; CAA, cerebral amyloid angiopathy;CACNA1A, P/Q-type calcium channel a1A subunit;CD, circular dichroism; Cdc, cell division cycle;Cdk 5, cyclin dependent kinase 5; ChAT, cholineacetyltransferase;CH-plot, charge-hydropathy plot; CJD, Creutzfeldt-Jakob disease; CLSM, confocal laser scanning microscopy;CS, Cockayne syndrome;CWD, chronic wasting disease;DLB, dementia with Lewy bodies;DLS, dynamic light scattering;DRPLA, dentatorubral-pallidoluysian atrophy; EM, electron microscopy;EPR, electron paramagnetic resonance spectroscopy; ER, endoplasmic reticulum;ERCC-6, excision repair cross-complementing rodent repair deficiency, complementation group 6;ERD, ER exit signal;ESEM, environmental scanning electron microscopy;FACT, facilitates chromatin transcription;FBD, familial British dementia;FDD, familial Danish dementia;FFI, fatal familial insomnia; FRET, fluorescence resonance energy transfer;FTIR, Fourier-transform infrared spectroscopy; FTLD, frontotemporal lobar degeneration;FUS, fused in sarcoma; G3BP, Ras-GTPase-activating protein SH3-domain-binding protein;GCIs, glial cytoplasmic inclusions;GdmCl, guanidinium chloride;GFAP, glial fibrillary acidic protein;GSK-3ß, glycogen synthase kinase 3 beta;GSS, Gerstmann-Sträussler-Scheinker; HAD, HIV-associated dementia; HCHWA-D, hereditary cerebral hemorrhage with amyloidosis-Dutch type; HD, Huntington's disease; HIV-1, human immunodeficiency virus-1;HSD, Hallervorden-Spatz disease; HYPK, huntingtin yeast-two hybrid protein K;IBM, inclusion body myopathy with early-onset Paget disease and frontotemporal dementia;IDP, intrinsically disordered protein; IDR, intrinsically disordered region; LB, Lewy body; LN, Lewy neurite; Lsm, like Sm;LsmAD, Lsm-associated domain; MAPK, microtubule associated protein kinase;MJD, Machado-Joseph disease; mRNP, messenger ribonucleoprotein;MSA, multiple system atrophy; MTDD, multi-targeted designed drug; mTORC1, mammalian target of rapamycin complex 1; CREB-regulated transcription coactivator 1; MWI, mechanically weak interactions; NAC, non-amyloid component; NACP, non-amyloid component precursor protein; NBIA1, neurodegeneration with brain iron accumulation type 1; NGF, nerve growth factor; NLS, nuclear localization signal; NFT, neurofibrillary tangle; ODR, optical rotary dispersion; OM, optical microscopy; PD, Parkinson's disease; PHF, paired helical filament; POAG, primary open-angle glaucoma; POLG, polymerase g; polyQ, polyglutamine repeat; PONDR, predictor of naturally disordered regions; proNGF, NGF precursor, without signal peptide; PrP, prion protein; PSD, p300 region similar to prion-like domains; p75NTR, pan-neurotrophin receptor; RF, Rosenthal fiber; RRM, RNA recognition motif; SAXS, small angle X-ray scattering; SANS, symmetrically dimethylated arginine residues; SDSI, site-directed spin labeling; SEPT, gene encoding septin; SG, stress granule; SMA, spinal muscular atrophy; SMN, survival of motor neurons; SMFS, single-molecule force spectroscopy; STAGA, SPT3-TAF9-ADA-GCN5 acetyltransferase; STM, scanning tunneling microscopy; SWI/SNF, SWItch/Sucrose NonFermentable; TBP, TATA-box-binding protein; TDP-43, TAR DNA-binding protein of 43 kDa; TIA-1, T cell internal antigen-1; TIAR, TIA-1-related; TIRFM, total internal reflection fluorescence microscopy; TTPP, tubulin polymerization promoting protein; TrkA, tropomyosin-related kinase A; TSE, transmissible spongiform encephalopathy; VCP, valosin-containing protein
Protein intrinsic disorder in degenerative diseases

Key Words: Neurodegenerative disease; intrinsically disordered protein; protein folding; protein misfolding; protein aggregation; protein-protein interaction; protein function; protein dysfunction

Send correspondence to: Vladimir N. Uversky, Department of Molecular Medicine, University of South Florida, Morsani College of Medicine, 12901 Bruce B. Downes Blvd, MDC07, Tampa, Florida 33612, Tel: 1-813-974-5816, Fax: 1-813-974-7357, E-mail: vuversky@health.usf.edu