Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Function and structure of inherently disordered proteins

A Keith Dunker[1], Israel Silman[2], Vladimir N Uversky[1,3,4] and Joel L Sussman[5]

The application of bioinformatics methodologies to proteins inherently lacking 3D structure has brought increased attention to these macromolecules. Here topics concerning these proteins are discussed, including their prediction from amino acid sequence, their enrichment in eukaryotes compared to prokaryotes, their more rapid evolution compared to structured proteins, their organization into specific groups, their structural preferences, their half-lives in cells, their contributions to signaling diversity (via high contents of multiple-partner binding sites, post-translational modifications, and alternative splicing), their distinct functional repertoire compared to that of structured proteins, and their involvement in diseases.

Addresses

[1] Center for Computational Biology and Bioinformatics, Institute for Intrinsically Disordered Protein Research, Indiana University Schools of Medicine and Informatics, Indianapolis, IN 46202, USA

[2] Department of Neurobiology and The Israel Structural Proteomics Center, Weizmann Institute of Science, Rehovot 76100, Israel

[3] Institute for Intrinsically Disordered Protein Research, Center for Computational Biology and Bioinformatics, Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN 46202, USA

[4] Institute for Biological Instrumentation, Russian Academy of Sciences, 142290 Pushchino, Moscow Region, Russia

[5] Department of Structural Biology and The Israel Structural Proteomics Center, Weizmann Institute of Science, Rehovot 76100, Israel

Corresponding author: Sussman, Joel L (Joel.Sussman@weizmann.ac.il)

Current Opinion in Structural Biology 2008, **18**:756–764

This review comes from a themed issue on Proteins
Edited by Janet Smith and Martino Bolognesi

Available online 17th November 2008

0959-440X/$ – see front matter
© 2008 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.sbi.2008.10.002

Introduction

Many entire proteins and localized protein regions fail to fold into a 3D structure, yet carry out function. Rather than a linear sequence-to-structure-to-function paradigm, such proteins have been described by a trinity in which function arises from different forms (structured globules, collapsed disordered ensembles, and extended disordered ensembles) and from transitions between different forms such as a disorder-to-structure transition upon binding [1]. The collapsed disordered ensembles were originally thought to be exclusively native molten globules (MGs), with collapse driven by hydrophobic interactions. Recent studies, however, agree with earlier work suggesting that water is a poor solvent for the peptide backbone; thus, polar but uncharged model sequences form compact random coils [2••,3••], while extended random coils result when polypeptide chains contain significant net charge (Rohit Pappu, unpublished). Water being a poor solvent for polypeptides was also invoked to explain the occurrence of a fourth protein form, the pre-MG, that occurs as an intermediate between the MG and the random coil during protein unfolding [4]. Much more work is needed to understand and relate the various nonstructured protein ensembles and to determine whether the relationship between structure and function should be assembled into a trinity, a quartet, or an even more complicated arrangement.

Here we provide an overview of these proteins, including their structures, functions, and regulations. In all these aspects, the set of non-folding proteins and regions is found to differ greatly from the set of proteins that fold into globular 3D structures.

Prediction of non-folding proteins and regions

Since the amino acid sequence contains the information for protein folding, it was reasoned that, for proteins that do not fold into 3D structures, the amino acid sequence should also specify protein non-folding. To test this hypothesis, predictors were developed to identify sequences that fail to fold [5,6]. The fact that predictor accuracy was significantly better than expected by chance suggested that the information for failure to fold into a 3D structure is, indeed, likely to be inherent within the amino acid sequence.

Compared to structured proteins, non-folding proteins are significantly depleted in C, W, Y, F, I, V, and L and significantly enriched in M, K, R, S, Q, P, and E. The reduced level of residues forming the hydrophobic cores of structured proteins, and the increased level of residues forming the surfaces of structured proteins, makes their failure to fold into a 3D structure easy to understand [7]. By now more than 20 predictors of protein disorder have been developed [7], including meta-predictors combining several methods [8]. A very useful overview of current predictors, especially with regard to providing help with protein crystalization, recently became available [9].

An interesting development has been the inclusion of disorder prediction in the biannual critical assessment of structure prediction (CASP) exercise [10•]. This has

stimulated the development of disorder predictors, and permitted comparison of various predictors in blind tests. The main finding at the most recent meeting, CASP7, was that, of 19 different predictors, most gave fairly good results. The best five reached overall accuracies of ~69% to ~78% for a two-state prediction, and also achieved areas under their receiver operating characteristics (ROC) curves 0.822 ± 0.008 to 0.860 ± 0.007 [10•], where a ROC curve is a graphical method of assessing predictor performance based on plotting its true positive rate versus its false positive rate. A random predictor would give a value of 0.5 for the area under the ROC curve, and a perfect predictor would give 1.0. Thus, the observed values, which are >0.80, indicate fairly good predictors.

Noise in both structured and disordered training data is most probably a major contributor to the limitations in these and other disorder predictors. If so, reducing the noise in the training data will be needed to obtain significant improvement in prediction accuracy.

Frequency of disordered regions

Disorder predictions have been carried out for many whole proteomes. They indicate that the fraction of proteins with substantial amounts of disorder goes as eukaryotes $\gg$ archaea ~ eubacteria, with multicellular eukaryotes having much more predicted disorder than mono-cellular eukaryotes [11]. These results were confirmed and substantially extended to include functional classification using an improved predictor of disorder [12]. Integrating the results from these and other sources gives some rules of thumb: for mammals, ~75% of their signaling proteins are predicted to contain long disordered regions (>30 residues), about half of their total proteins are predicted to contain such long disordered regions, and ~25% of their proteins are predicted to be fully disordered.

Protein evolution

Non-folding proteins and regions might be expected to change more rapidly during evolution than structured proteins because buried amino acids are highly constrained while disordered regions are not constrained by structure. For example, plots of sequence variability (measured by sequence entropy over alignments) were found to exhibit nearly linear dependence on the inverse of the packing density, until a low packing density was reached at which point sequence variability remained roughly constant as packing density decreased further [13]. Equivalent results were found in studies showing that amino acid changes are more rapid for proteins with greater proportions of solvent-exposed amino acids [14]. Against this background, non-folding regions of a given protein would be expected to exhibit lower sequence conservation relative to structured regions in the same protein, and this appears to hold for most proteins studied [15].

Structural and dynamical examination using NMR shows that the ~70-residue intrinsically disordered linker in replication protein A maintains essentially the same length, and a similar degree of flexibility, across widely divergent species, even though the various sequences show almost no measurable sequence identity [16••]. Thus, for disordered regions with linker function, conserving length and composition may be more important than sequence conservation for maintaining function. Indeed, no obvious correlations could be observed between the rate of accumulated changes and the types of protein function for the various regions of disorder [15]. The relationships among the sequence of a disordered region, its rate of evolution, and its function thus merit additional, concerted investigation.

Partitioning unstructured proteins and regions into groups

Grouping proteins according to structure and function has proven very useful for studying structured proteins. Associating a new protein with an existing structure–function group (by sequence and/or structure alignment) provides important basic information and quickly identifies critical experiments for further characterization. Given the broad array of disordered protein types, their lack of 3D structure, and their sequence variability, it has so far proven difficult to cluster various non-folding proteins and regions into distinct groups to provide the basis sets for the analysis of new disordered proteins and regions.

Functional disordered regions can be as short as a few residues, such as the four to eight amino acid linker in calmodulin that forms a helix in the crystal structure but is flexible in solution. This flexible linker works together with its flanking calcium-binding domains to surround their binding target, which is typically a short helical segment. The disorder of the linker and the flexibility of the side-chains in the methionine-rich hydrophobic patches of the calcium-binding regions enable calmodulin to bind to a wide variety of target sequences [17]. In contrast, the PEVK disordered region of titin ranges in length from 180 residues in the cardiac muscle isoform to 2174 residues in the soleus muscle isoform. These regions function as entropic springs to assist maintenance of the appropriate length of the muscle fibers [18]. Thus, disordered regions exhibit a broad continuum of lengths, display a wide variety of functions, and show wide differences in their tendencies to form structures [19]. All these factors suggest that clustering disordered proteins into groups will not be simple.

Following a number of failed attempts to partition a collection of disordered sequences into groups by sequence analysis methods (Dunker, Brown, unpublished), an attempt was made to group disordered regions into homogeneous subsets by an alternative approach. Briefly, disordered regions were randomly partitioned

into subsets, different predictors were constructed for each subset, the collections of disordered regions were repartitioned into groups according to which predictor gave the best results, new predictors were developed on the basis of the repartitioned subsets, and the steps were repeated until the predictions and partitions no longer changed from one cycle to the next. By this approach three different types, or flavors, of disordered regions were found, and these were arbitrarily called V, C, and S [20].

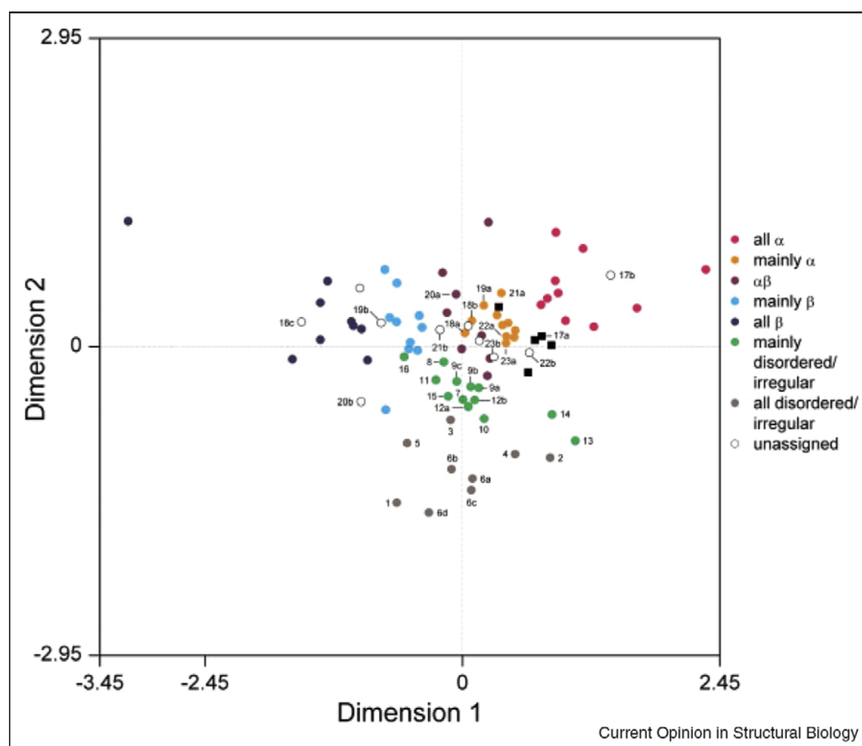
Interestingly, some functional partitioning was observed among the three flavors, with flavor S including many examples of protein-binding regions, flavor V being rich in ribosomal proteins, and flavor C containing a higher number of sites of protein modification [20]. Despite the anticipated difficulties, new attempts to cluster disordered proteins and disordered regions into subsets should be undertaken.

Do inherently unstructured proteins retain any preference for certain structures or are they totally unstructured?

One of the key open questions regarding inherently unstructured proteins is whether, in solution, they retain some preferred structure(s), or are just a plethora of many different conformations, rather like 'cooked spaghetti'. A recent careful study of residual structure in disordered peptides and unfolded proteins was carried out via multivariate analysis and *ab initio* simulation of Raman optical activity [21•]. This study showed striking differences between the structural characteristics of natively unfolded proteins and proteins unfolded by denaturation. The former tend to cluster in the mainly disordered/irregular region of the nonlinear mapping plot of Figure 1, and seem to contain a significant amount of the extended PPII-helical conformation; the latter appear in other regions and can contain significant amounts of β structure in the case of reduced proteins, and of α-helix in the case of acid MGs.

Human neuroligin 3 (hNL3) is a member of the family of cholinesterase-like adhesion molecules (CLAMs). It is a single-pass transmembrane protein, and preliminary bioinformatics analysis indicated that like the cytoplasmic domain of the *Drosophila* CLAM, gliotactin, its cytoplasmic domain (hNL3-cyt) is intrinsically disordered [22]. A detailed study of hNL3-cyt using sensitivity to proteases and a variety of biophysical and spectroscopic methods, showed that it is, indeed, intrinsically unstructured [23], as bioinformatics analysis had suggested

Figure 1



Two-dimensional nonlinear mapping plot of a set of 85 peptide, protein and virus Raman optical activity spectra [21•]. More complete definitions of the structure classes are: all α, >60% α-helix with little or no β-sheet; mainly α, >35% α-helix, and a small amount (<15%) or no β-sheet; αβ, significant amounts (>15%) of both α-helix and β-sheet; mainly β, >35% β-sheet, and a small amount (<15%) or no α-helix; all β, >45% β-sheet with little or no α-helix; mainly disordered/irregular, little α-helix or β-sheet; all disordered/irregular, no α-helix or β-sheet. The black squares identify a subset of mainly α proteins containing mostly α-helix and disordered structure with little or no β-sheet.

[22,24]. However, some biophysical techniques indicated that hNL3-cyt is not fully extended, but becomes significantly more so under denaturing conditions, implying residual structure in this largely unstructured protein sequence. This increased extension is in agreement with the proposition that raising the denaturant concentration produces increased polyproline II content in short peptides and proteins [25]. While denaturation of a structured protein can be considered as a transition from a globally structured state to a dynamic ensemble of locally structured states, denaturation of many natively unfolded proteins cannot be viewed in such terms. Rather, their denaturation can be seen as a transition from a set of globally flexible, extended, and largely unstructured states to a dynamic ensemble of even more extended unstructured states [23].

Do non-folding proteins have a shorter half-life than other proteins?

Targeted turnover of proteins is a key element in the regulation of many cellular processes. The underlying physicochemical and/or sequential signals are not, however, fully understood. This is particularly pertinent in light of recent recognition that intrinsically unstructured/disordered proteins, common in eukaryotic cells, are extremely susceptible to proteolytic degradation *in vitro*. An *in vivo* high-throughput study of the half-lives of all yeast gene products [26] indicated that, in general, a protein's half-life does not depend on the presence of degradation signals within its sequence, even of ubiquitination sites, but correlates mainly with the length of its polypeptide chain and with various measures of structural disorder [27•]. The significance of correlation is rather low; thus protein degradation is not determined by a single characteristic, but is a multifactorial process displaying large protein-to-protein variations. Protein disorder, nevertheless, plays a crucial signaling role in many cases. This may be related to the capacity of the 20S proteasome to distinguish between intrinsically unstructured and other proteins. Specifically, it has recently been shown that 20S proteasomes can digest intrinsically unstructured sequences under conditions in which native, and even MG states, are resistant to degradation [28•].

Functionality of inherently disordered proteins and regions

Non-folding proteins and regions carry out pivotal biological functions, participating in various signaling and regulatory pathways, via specific protein–protein, protein–nucleic acid, and protein–ligand interactions [29–32]. Enzymatically controlled sites of post-translational modification (PTM) such as acetylation, hydroxylation, ubiquitination, methylation, and phosphorylation, as well as sites of proteolytic attack, are frequently associated with regions of intrinsic disorder [29]. Specific functions of nonstructured proteins and regions can be grouped into four broad classes: first, molecular recognition; second, molecular assembly; third, protein modification; fourth, entropic chain activities [33]. The capability of non-folding proteins and regions to interact with collections of partners is utilized in organizing complex protein–protein interaction networks. In fact, hub proteins were shown to have multiple interactions, either being intrinsically disordered and serving as an anchor, or acting as a stable globular scaffold that interacts with intrinsically disordered regions of its targets [34–40].

Recently, the disorder-associated and structure-associated functions in Swiss-Prot were identified [29–31]. Out of 710 functional keywords, each associated with at least 20 proteins, 310 were found to be structure-associated, 238 disorder-associated, and 170 structurally ambiguous [29–31]. The functional diversity provided by non-folding proteins and regions was found to complement the functions of structured proteins [29–31]. When functional keywords were grouped into 11 gene ontology categories, order-associated functions fell into only seven, while disorder-associated functions covered essentially all functional categories [29]. This implies that the functional repertoire of unfolded proteins and regions is larger than that of structured proteins [41]. Overall, structured proteins were mainly found to be associated with catalysis and transport, while non-folding proteins and regions were involved in signaling and regulatory processes [29–31].

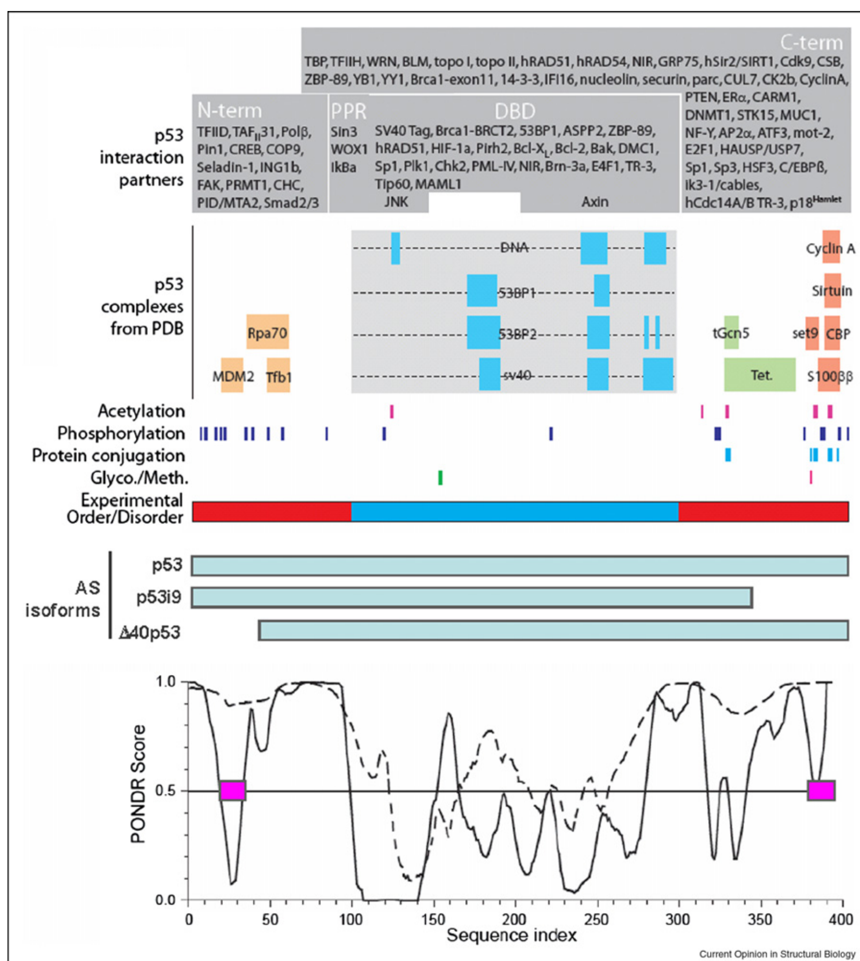
Nonstructured proteins and regions are often involved in molecular interactions mediated by localized binding sites identified by various terms such as eukaryotic linear motifs (ELMs), short linear motifs (SLiMs), and molecular recognition features (MoRFs). ELMs [42] and SLiMs [43] are both identified as short sequence patterns in multiple proteins that bind to a common target, while MoRFs are identified by a pattern in a disorder prediction output [44]. Despite being detected by different methods, ELMs and SLiMs have high overlap with MoRFs [45•]. While α-MoRFs can be predicted with high accuracy [46], ELMs and SLiMs, with their short (<10 amino acid residue) and degenerate sequences, are more difficult to identify. SLiMs in non-folding regions exhibit greater sequence conservation than their flanking sequences, while SLiMs in structured regions exhibit similar sequence conservation compared to their flanking regions. This suggests that SLiMs in disordered regions are more likely to be biologically relevant [47].

Besides being placeholders for specific binding sites, non-folding regions are also the primary loci of alternative splicing (AS) [48] and of various PTMs [31]. Studies on the tumor suppressor, p53, show how these factors can interact to result in signaling diversity. The spectrum of p53-based cell fate decisions ranges from transient cell-cycle arrest through irreversible block of proliferation, to induction of senescence, differentiation, or apoptosis. The p53 molecule, being a multidomain protein at the

center of a large signaling network, interacts with DNA and with many proteins to carry out its various signal transduction functions [35,49]. The p53 DNA-binding domain is structured, whereas the N-terminal and C-terminal domains are disordered (reviewed in [35,49]). Figure 2 shows that 60/84 = 71% of p53 interactions are mediated by the non-folding regions that comprise only ~29% of its sequence. p53 has multiple PTMs (Figure 2, vertical ticks), with 86%, 90%, and 100% of observed acetylation, phosphorylation, and protein conjugation sites, respectively, located in the disordered regions. α-MoRFs are predicted in both the transactivation domain and the regulatory domain (Figure 2, pink bars), and these two domains also contain most of the PTM sites [35]. Analysis of the ability of the C-terminal α-MoRF to bind to multiple, structurally distinct protein partners showed that extreme structural plasticity of the disordered region is important for binding to the structurally distinct partners [35].

There are three major p53 AS isoforms: full-length p53, C-terminally truncated p53i9, lacking the last 60 residues, and N-terminally truncated Δ40p53, which is depleted of the first 40 residues [50]. The Δ40p53 isoform inhibits p53 transcriptional activity and p53-mediated apoptosis, modifies p53 cellular localization, and inhibits Mdm2-mediated p53 degradation [50]. The p53i9 isoform is defective in transcriptional activity and devoid of

Figure 2



Summary of p53 interactions and structure. Gray boxes indicate the approximate binding regions of its known binding partners. The regions of p53 cocrystalized with binding partners and represented in Protein Data Bank structures of corresponding complexes are shown by tan, blue, green, and peach horizontal bars, each labeled with the name of the binding partner. The light gray box corresponds to the structured DNA-binding domain. The blue bars within it identify regions directly involved in interaction with a given binding partner, whose name is shown at the middle of the corresponding dashed line. PTM sites are represented by colored vertical ticks. Experimentally characterized regions of disorder (red) and order (blue) are indicated by the horizontal bar. Alternatively spliced p53 isoforms are shown by gray–blue horizontal bars. Finally, predictions of disorder (scores >0.5) and order (scores <0.5) are shown for two PONDR predictors: VLXT (solid line) and VSL2P (dashed line). Predicted α-MoRFs are shown by pink bars. All features are presented to scale, as indicated by the horizontal axis. The p53 interaction partners and PTM sites are adapted from Anderson and Appella [51] (Figure modified from [35]).

DNA-binding activity [50]. In support of these observations, Figure 2 shows that AS eliminates substantial portions of the two terminal disordered regions, thereby removing the α-MoRFs, numerous PTMs, and multiple additional protein-binding sites. As exemplified by p53, we believe that the convergence of AS sites, PTMs, and various binding sites, all in disordered regions, makes these intrinsically unstructured regions focal points for enhanced signaling diversity.

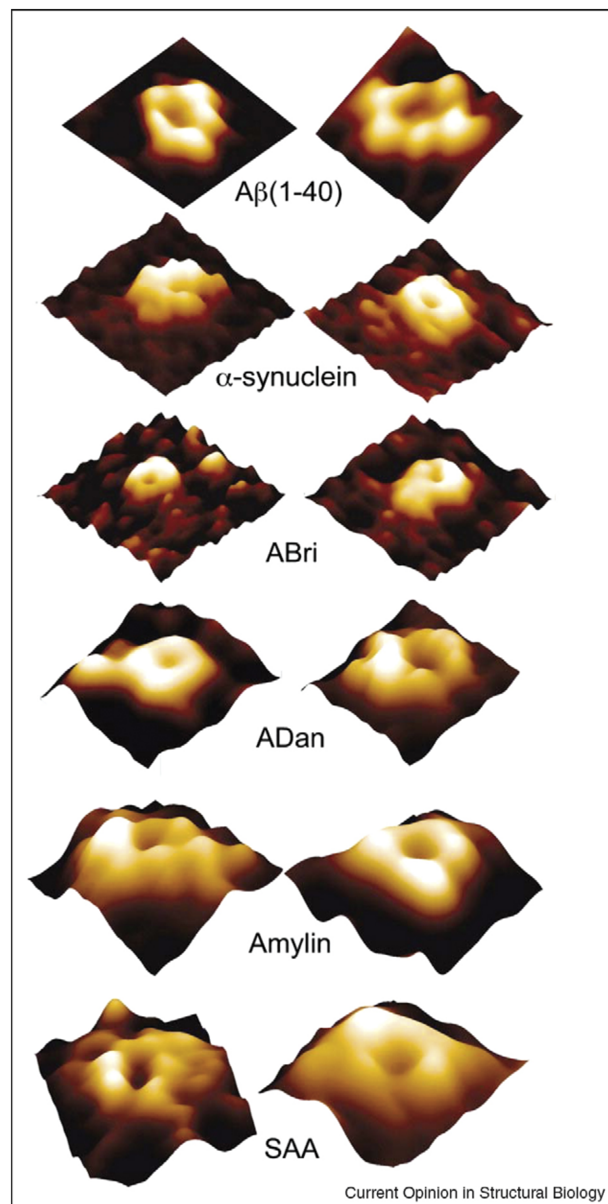
Involvement of inherently disordered proteins in diseases

The fact that many proteins are either wholly intrinsically disordered, or contain large stretches of intrinsically disordered sequences, has been followed by a growing realization that nonstructured proteins are associated with a broad range of human diseases, which led to the introduction of the D^2 (disorder in disorders) concept [52]. Diseases involving protein disorder come in a variety of flavors, but we here restrict ourselves to discussing recent work concerning the amyloid diseases, in which the aggregation of misfolded protein sequences to amyloid fibrils rich in β-structure, and, eventually, into larger entities, is associated with the pathogenesis of some of the most prominent neurodegenerative diseases. These include Alzheimer's disease (AD), Parkinson's disease (PD), Huntington's disease, and the prion diseases [53].

Although a circumstantial case can be made that amyloid fibril formation is a primary event in the disease process [54••], it has been realized for some time, in particular for AD, that the levels of Aβ deposits detected postmortem do not correlate well with the degree of severity of the disease [55]. Consequently, it has been suggested that oligomers or protofibrils of the initially disordered polypeptides are the pathogenic entities involved, and that their mode of action may involve formation of pores in the plasma membranes of the cells affected (reviewed by [54••]). One of the earliest studies to raise this possibility reported production by Aβ(1–40) of large cation channels in planar lipid bilayers [56]; subsequently Aβ(1–40) was reported to generate Zn^{2+}-sensitive Ca^{2+}-channels in reconstituted liposomes [57]. Quist *et al.* [58•] used various techniques to show that amyloid peptides associated with several diseases produced similar channels, and utilized atomic force microscopy to visualize pore-like structures for amyloid peptides reconstituted into lipid bilayers (Figure 3). In the review cited above, Lashuel and Lansbury [54••] contended that the amyloid peptides mimicked the action of bacterial pore-forming toxins, and suggested that they might thus form oligomeric membrane-spanning pores with β-barrel structures.

The synucleins are a family of three highly homologous proteins, α-synuclein (αSyn), β-synuclein (βSyn), and γ-synuclein (γSyn), containing ∼130 amino acid residues, which are typical intrinsically disordered proteins [59]. The aggregation of αSyn into oligomers, protofibrils, and fibrils is closely related to the pathogenesis of PD, Lewy body dementia, and several other neurodegenerative diseases known as synucleinopathies [60,61]. Despite the similarity in their biophysical properties, both βSyn and γSyn are less prone to fibrillate than αSyn, and both inhibit fibril formation by αSyn [59].

Figure 3



Individual channel-like structures formed by amyloid polypeptides visualized by AFM. Two examples are shown for each molecule, in which a central pore can be observed. [Image sizes are 25 nm for Aβ(1–40), 25 nm for β-synuclein, 35 nm for ABri, 20 nm for ADan, 25 nm for amylin, and 20 nm for SAA.] Figure taken from Ref. [58•].

In general, αSyn has been shown to possess remarkable structural plasticity, being capable of adopting a series of structurally unrelated conformations, whose features are strongly dependent on the protein environment and/or on the available binding partners [62]. There is extensive literature documenting assembly of αSyn into fibrils with high β-sheet content [63–65]. However, it has also been reported that it assumes an α-helical conformation when associated with phospholipids [66,67] or with detergent micelles [65,68].

A recent study [69•] examined the physicochemical characteristics of αSyn bound to liposomes and to planar bilayers. It showed that whereas αSyn in solution had a far-UV-CD spectrum of a completely unordered protein, when bound to liposomes containing anionic phospholipids it adopted a predominantly α-helical conformation. Upon addition to a planar bilayer membrane composed of anionic phospholipids, monomeric αSyn produced discrete ion channels. The authors inferred that the ion channels observed were produced by one or several molecules of αSyn in an α-helical conformation, and that they might be involved in its normal function and/or pathophysiology.

Another recent study [70••] used both theoretical and experimental approaches to examine the aggregation and pore-forming capacity of αSyn. Molecular modeling and molecular dynamics (MDs) predicted that αSyn adopts both nonpropagating and propagating conformations, with the latter leading to ring-like pentameric and hexameric structures. Analogous simulations on βSyn did not produce ring-like structures. Furthermore, they showed that βSyn can bind to αSyn, creating stable nonpropagating heterodimers. In their experiments, the authors demonstrated, using negative staining, formation of ring-like structures resembling those predicted by the modeling studies, and further demonstrated their gradual disappearance upon addition of βSyn. They went on to show appearance of nonselective ion channels in HEK293 cells overexpressing αSyn, which were inhibited by Zn^{2+} similarly to those elicited by Aβ(1–40) described earlier [57]. Furthermore, they were not detected in cells coexpressing αSyn and βSyn. If the pore-forming activity of αSyn is associated with the neurodegenerative process, rather than with its innate physiological activity, the ability of βSyn to attenuate pore formation may help explain the protective effects of βSyn reported both *in vitro* and *in vivo* [71]; these, in turn, may help to develop therapeutic strategies for reducing αSyn aggregation in PD and related disorders.

Conclusions

The sequence-to-structure-to-function paradigm for proteins was developed from the study of enzymes. Bioinformatics studies indicate that this paradigm applies to enzymes, as well as to transport proteins.

In contrast, proteins and regions of proteins involved in signaling, control, and regulation often use inherently unstructured sequences as the basis for function. There are many structured signaling domains, but these often bind to unstructured protein partners. Moreover, there are numerous signaling enzymes, such as kinases, acetylases, and methylases, but these typically act on unfolded protein sequences. Compared to structured proteins, flexible, unfolded protein ensembles apparently have fundamental advantages for signaling by being more capable of responding to changes in their environments.

The use of bioinformatics methods of prediction and classification has greatly enhanced research on intrinsically unstructured proteins. Before the application of these methods, many reports had described biological functions being carried out by flexible, unstructured proteins, and regions. For more than 40 years, however, these various studies remained unconnected. Bioinformatics identified commonalities among large numbers of these proteins, bioinformatics showed that they are common, not rare, and bioinformatics identified the wide-spread use of such proteins and regions for signaling and regulation.

Acknowledgements

This work was supported in part by the grants R01 LM007688-01A1 (to AKD and VNU) and GM071714-01A2 (to AKD and VNU) from the National Institutes of Health and from the Program of the Russian Academy of Sciences 'Molecular and Cellular Biology' (to VNU), by the Divadol Foundation, the Benoziyo Center for Neuroscience, the Kimmelman Center, Autism Speaks, the Israel Science Foundation, the Nalvyco Foundation, the Neuman Foundation, a research grant from Mr. Erwin Pearl, the European Commission Sixth Framework Research and Technological Development Programme 'SPINE2-COMPLEXES' Project under contract number LSHG-CT-2006-031220 and 'Teach-SG' Project, under contract number ISSG-CT-2007-037198 (IS and JLS). JLS is the Morton and Gladys Pickman Professor of Structural Biology. We gratefully acknowledge the support of the IUPUI Signature Centers Initiative.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Dunker AK, Obradovic Z: **The protein trinity — linking function and disorder**. *Nat Biotechnol* 2001, **19**:805-806.

2. •• Crick SL, Jayaraman M, Frieden C, Wetzel R, Pappu RV: **Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions**. *Proc Natl Acad Sci U S A* 2006, **103**:16764-16769.
See annotation to [3••].

3. •• Tran HT, Mao A, Pappu RV: **Role of backbone-solvent interactions in determining conformational equilibria of intrinsically disordered proteins**. *J Am Chem Soc* 2008, **130**:7380-7392.
Along with [2••], this paper used theoretical predictions and experiments to test the prediction that highly polar, uncharged sequences collapse despite their lack of hydrophobic residues. This collapse occurs for a variety of polar uncharged sequences and even for the backbone alone. Thus, it is likely to be a general result that arises because water is such a poor solvent for both the backbone and certain polar polypeptide chains.

4. Uversky VN: **Protein folding revisited. A polypeptide chain at the folding–misfolding–nonfolding cross-roads: which way to go?** *Cell Mol Life Sci* 2003, **60**:1852-1871.

5. Romero P, Obradovic Z, Kissinger C, Villafrance JE, Dunker AK: **Identifying disordered regions in proteins from amino acid sequence**. In *Proceedings of the International Conference on Neural Networks*. 1997:91-95.

6. Uversky VN, Gillespie JR, Fink AL: **Why are "natively unfolded" proteins unstructured under physiologic conditions?** *Proteins* 2000, **41**:415-427.

7. Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK: **Intrinsic disorder and functional proteomics**. *Biophys J* 2007, **92**:1439-1456.

8. Ishida T, Kinoshita K: **Prediction of disordered regions in proteins based on the meta approach**. *Bioinformatics* 2008, **24**:1344-1348.

9. Dosztanyi Z, Tompa P: **Prediction of protein disorder**. *Methods Mol Biol* 2008, **426**:103-115.

10. • Bordoli L, Kiefer F, Schwede T: **Assessment of disorder predictions in CASP7**. *Proteins* 2007, **69(Suppl 8)**:129-136.

The CASP assessment papers are worth reading. Various researchers claim that one predictor or another (usually their own) is better than the others. The CASP exercises are evaluations based on blind predictions, and the evaluations are by third parties, so the comparisons are less biased. However, the CASP exercises involve only a small number of proteins, so claims to have developed 'the best predictor' should not be made on their basis either.

11. Dunker AK, Obradovic Z, Romero P, Garner EC, Brown CJ: **Intrinsic protein disorder in complete genomes**. *Genome Inform Ser Workshop Genome Inform* 2000, **11**:161-171.

12. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT: **Prediction and functional analysis of native disorder in proteins from the three kingdoms of life**. *J Mol Biol* 2004, **337**:635-645.

13. Jernigan RL, Kloczkowski A: **Packing regularities in biological structures relate to their dynamics**. *Methods Mol Biol* 2007, **350**:251-276.

14. Lin YS, Hsu WL, Hwang JK, Li WH: **Proportion of solvent-exposed amino acids in a protein and rate of protein evolution**. *Mol Biol Evol* 2007, **24**:1005-1011.

15. Brown CJ, Takayama S, Campen AM, Vise P, Marshall TW, Oldfield CJ, Williams CJ, Dunker AK: **Evolutionary rate heterogeneity in proteins with long disordered regions**. *J Mol Evol* 2002, **55**:104-110.

16. •• Daughdrill GW, Narayanaswami P, Gilmore SH, Belczyk A, Brown CJ: **Dynamic behavior of an intrinsically unstructured linker domain is conserved in the face of negligible amino acid sequence conservation**. *J Mol Evol* 2007, **65**:277-288.

This paper reminds us that not all functions associated with disorder involve binding to a partner, and shows that features likely to be of functional importance, such as length and flexibility, can be conserved in the face of lack of amino acid sequence conservation.

17. Fiorin G, Biekofsky RR, Pastore A, Carloni P: **Unwinding the helical linker of calcium-loaded calmodulin: a molecular dynamics study**. *Proteins* 2005, **61**:829-839.

18. Nagy A, Grama L, Huber T, Bianco P, Trombitas K, Granzier HL, Kellermayer MS: **Hierarchical extensibility in the PEVK domain of skeletal-muscle titin**. *Biophys J* 2005, **89**:329-336.

19. Dyson HJ, Wright PE: **Intrinsically unstructured proteins and their functions**. *Nat Rev Mol Cell Biol* 2005, **6**:197-208.

20. Vucetic S, Brown CJ, Dunker AK, Obradovic Z: **Flavors of protein disorder**. *Proteins* 2003, **52**:573-584.

21. • Zhu F, Kapitan J, Tranter GE, Pudney PD, Isaacs NW, Hecht L, Barron LD: **Residual structure in disordered peptides and unfolded proteins from multivariate analysis and ab initio simulation of Raman optical activity data**. *Proteins* 2008, **70**:823-833.

This paper shows the power of Raman optical activity to distinguish between intrinsically unfolded sequences and sequences that fold into structures that are often difficult to differentiate from them, such as the polyproline-II fold.

22. Zeev-Ben-Mordehai T, Rydberg EH, Solomon A, Toker L, Botti S, Auld VJ, Silman I, Sussman JL: **The intracellular domain of the *Drosophila* cholinesterase-like neural adhesion protein, gliotactin, is natively unfolded**. *Proteins* 2003, **53**:758-767.

23. Paz A, Zeev-Ben-Mordehai T, Lundqvist M, Sherman E, Mylonas E, Weiner L, Haran G, Svergun D, Mulder F, Sussman JL *et al.*: **Biophysical characterization of the unstructured cytoplasmic domain of the human neuronal adhesion protein Neuroligin 3**. *Biophys J* 2008, **95**:1928-1944.

24. Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg E, Man O, Beckmann JS, Silman I, Sussman JL: **FoldIndex©: a simple tool to predict whether a given protein sequence is intrinsically unfolded**. *Bioinformatics* 2005, **21**:3435-3438.

25. Whittington SJ, Chellgren BW, Hermann VM, Creamer TP: **Urea promotes polyproline II helix formation: implications for protein denatured states**. *Biochemistry* 2005, **44**:6269-6275.

26. Belle A, Tanay A, Bitincka L, Shamir R, O'Shea EK: **Quantification of protein half-lives in the budding yeast proteome**. *Proc Natl Acad Sci U S A* 2006, **103**:13004-13009.

27. • Tompa P, Prilusky J, Silman I, Sussman JL: **Structural disorder serves as a weak signal for intracellular protein degradation**. *Proteins* 2008, **71**:903-909.

These results were a big surprise; most expected unfolded proteins to be degraded faster than structured proteins. Intracellular protein lifetimes vary over at least 12 orders of magnitude. Evidently a number of factors, including protein disorder, contribute to intracellular protein lifetimes.

28. • Tsvetkov P, Asher G, Paz A, Reuven N, Sussman JL, Silman I, Shaul Y: **Operational definition of intrinsically unstructured protein sequences based on susceptibility to the 20S proteasome**. *Proteins* 2008, **70**:1357-1366.

This paper gives a good operational definition, in solution, of intrinsically unstructured proteins, and provides a basis for distinguishing them, *in situ*, from MGs and from folded proteins.

29. Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Uversky VN, Obradovic Z: **Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions**. *J Proteome Res* 2007, **6**:1882-1898.

30. Vucetic S, Xie H, Iakoucheva LM, Oldfield CJ, Dunker AK, Obradovic Z, Uversky VN: **Functional anthology of intrinsic disorder. 2. Cellular components, domains, technical terms, developmental processes, and coding sequence diversities correlated with long disordered regions**. *J Proteome Res* 2007, **6**:1899-1916.

31. Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Obradovic Z, Uversky VN: **Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins**. *J Proteome Res* 2007, **6**:1917-1932.

32. Tompa P: **The interplay between structure and function in intrinsically unstructured proteins**. *FEBS Lett* 2005, **579**:3346-3354.

33. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z: **Intrinsic disorder and protein function**. *Biochemistry* 2002, **41**:6573-6582.

34. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN: **Flexible nets. The roles of intrinsic disorder in protein interaction networks**. *FEBS J* 2005, **272**:5129-5148.

35. Oldfield CJ, Meng J, Yang JY, Yang MQ, Uversky VN, Dunker AK: **Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners**. *BMC Genomics* 2008, **9(Suppl 1)**:S1.

36. Patil A, Nakamura H: **Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks**. *FEBS Lett* 2006, **580**:2041-2045.

37. Haynes C, Oldfield CJ, Ji F, Klitgord N, Cusick ME, Radivojac P, Uversky VN, Vidal M, Iakoucheva LM: **Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes**. *PLoS Comput Biol* 2006, **2**:e100.

38. Dosztanyi Z, Chen J, Dunker AK, Simon I, Tompa P: **Disorder and sequence repeats in hub proteins and their implications for network evolution**. *J Proteome Res* 2006, **5**:2985-2995.

39. Ekman D, Light S, Bjorklund AK, Elofsson A: **What properties characterize the hub proteins of the protein–protein interaction network of *Saccharomyces cerevisiae*?** *Genome Biol* 2006, **7**:R45.

40. Singh GP, Ganapathi M, Sandhu KS, Dash D: **Intrinsic unstructuredness and abundance of PEST motifs in eukaryotic proteomes**. *Proteins* 2006, **62**:309-315.

41. Daughdrill GW, Pielak GJ, Uversky VN, Cortese MS, Dunker AK: **Natively disordered proteins**. In *Handbook of Protein Folding.* Edited by Buchner J, Kiefhaber T. Verlag GmbH & Co: Wiley-VCH; 2005:271-353.

42. Puntervoll P, Linding R, Gemund C, Chabanis-Davidson S, Mattingsdal M, Cameron S, Martin DM, Ausiello G, Brannetti B, Costantini A *et al.*: **ELM server: a new resource for investigating short functional sites in modular eukaryotic proteins**. *Nucleic Acids Res* 2003, **31**:3625-3630.

43. Edwards RJ, Davey NE, Shields DC: **SLiMFinder: a probabilistic method for identifying over-represented, convergently evolved, short linear motifs in proteins**. *PLoS ONE* 2007, **2**:e967.

44. Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, Dunker AK: **Coupled folding and binding with alpha-helix-forming molecular recognition elements**. *Biochemistry* 2005, **44**:12454-12470.

45. • Fuxreiter M, Tompa P, Simon I: **Local structural disorder imparts plasticity on linear motifs**. *Bioinformatics* 2007, **23**:950-956.
Disordered regions commonly contain localized regions that bind to specific partners. This paper shows that different bioinformatics methods for identifying binding sites often point to the same or overlapping localized regions despite the large differences in the methods used.

46. Cheng Y, Oldfield CJ, Meng J, Romero P, Uversky VN, Dunker AK: **Mining alpha-helix-forming molecular recognition features with cross species sequence alignments**. *Biochemistry* 2007, **46**:13468-13477.

47. Ren S, Uversky VN, Chen Z, Dunker AK, Obradovic Z: **Short linear motifs recognized by SH2, SH3 and S/T kinase domains are conserved in disordered protein regions**. *BMC Genomics* 2008, **9**:S26.

48. Romero PR, Zaidi S, Fang YY, Uversky VN, Radivojac P, Oldfield CJ, Cortese MS, Sickmeier M, LeGall T, Obradovic Z *et al.*: **Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms**. *Proc Natl Acad Sci U S A* 2006, **103**:8390-8395.

49. Joerger AC, Fersht AR: **Structural biology of the tumor suppressor p53**. *Annu Rev Biochem* 2008, **77**:557-582.

50. Bourdon JC, Fernandes K, Murray-Zmijewski F, Liu G, Diot A, Xirodimas DP, Saville MK, Lane DP: **p53 isoforms can regulate p53 transcriptional activity**. *Genes Dev* 2005, **19**:2122-2137.

51. Anderson CW, Appella E: **Signaling to the p53 tumor suppressor through pathways activated by genotoxic and nongenotoxic stress**. In *Handbook of Cell Signaling.* Edited by Bradshaw RA, Dennis EA. Academic Press; 2003:237-247.

52. Uversky VN, Oldfield CJ, Dunker AK: **Intrinsically disordered proteins in human diseases: introducing the D2 concept**. *Annu Rev Bipohys Biomol Struct* 2008, **37**:215-246.

53. Uversky VN: **Amyloidogenesis of natively unfolded proteins**. *Curr Alzheimer Res* 2008, **5**:260-287.

54. •• Lashuel HA, Lansbury PT Jr: **Are amyloid diseases caused by protein aggregates that mimic bacterial pore-forming toxins?** *Quart Rev Biophys* 2006, **39**:167-201.
This is a critical review of the evidence that amyloid peptides cause disease by a mechanism involving pore formation.

55. Terry RD, Masliah E, Salmon DP, Butters N, DeTeresa R, Hill R, Hansen LA, Katzman R: **Physical basis of cognitive alterations in Alzheimer's disease: synapse loss is the major correlate of cognitive impairment**. *Ann Neurol* 1991, **30**:572-580.

56. Arispe N, Pollard HB, Rojas E: **Giant multilevel cation channels formed by Alzheimer disease amyloid beta-protein [A beta P-(1–40)] in bilayer membranes**. *Proc Natl Acad Sci U S A* 1993, **90**:10573-10577.

57. Lin H, Zhu YJ, Lal R: **Amyloid beta protein (1–40) forms calcium-permeable, Zn^{2+}-sensitive channel in reconstituted lipid vesicles**. *Biochemistry* 1999, **38**:11189-11196.

58. • Quist A, Doudevski I, Lin H, Azimova R, Ng D, Frangione B, Kagan B, Ghiso J, Lal R: **Amyloid ion channels: a common structural link for protein-misfolding disease**. *Proc Natl Acad Sci U S A* 2005, **102**:10427-10432.
In this study atomic force microscopy is used to show that several amyloid peptides form ion-channel-like structures in reconstituted membranes.

59. Uversky VN, Li J, Souillac P, Millett IS, Doniach S, Jakes R, Goedert M, Fink AL: **Biophysical properties of the synucleins and their propensities to fibrillate: inhibition of alpha-synuclein assembly by beta- and gamma-synucleins**. *J Biol Chem* 2002, **277**:11970-11978.

60. Galvin JE, Lee VM, Trojanowski JQ: **Synucleinopathies: clinical and pathological implications**. *Arch Neurol* 2001, **58**:186-190.

61. Uversky VN: **Neuropathology, biochemistry, and biophysics of alpha-synuclein aggregation**. *J Neurochem* 2007, **103**:17-37.

62. Uversky VN: **A protein-chameleon: conformational plasticity of alpha-synuclein, a disordered protein involved in neurodegenerative disorders**. *J Biomol Struct Dyn* 2003, **21**:211-234.

63. Serpell LC, Berriman J, Jakes R, Goedert M, Crowther RA: **Fiber diffraction of synthetic alpha-synuclein filaments shows amyloid-like cross-beta conformation**. *Proc Natl Acad Sci U S A* 2000, **97**:4897-4902.

64. Kim HY, Heise H, Fernandez CO, Baldus M, Zweckstetter M: **Correlation of amyloid fibril beta-structure with the unfolded state of alpha-synuclein**. *Chembiochem* 2007, **8**:1671-1674.

65. Rivers RC, Kumita JR, Tartaglia GG, Dedmon MM, Pawar A, Vendruscolo M, Dobson CM, Christodoulou J: **Molecular determinants of the aggregation behavior of alpha- and beta-synuclein**. *Protein Sci* 2008, **17**:887-898.

66. Davidson WS, Jonas A, Clayton DF, George JM: **Stabilization of alpha-synuclein secondary structure upon binding to synthetic membranes**. *J Biol Chem* 1998, **273**:9443-9449.

67. Zhu M, Fink AL: **Lipid binding inhibits alpha-synuclein fibril formation**. *J Biol Chem* 2003, **278**:16873-16877.

68. Ulmer TS, Bax A, Cole NB, Nussbaum RL: **Structure and dynamics of micelle-bound human alpha-synuclein**. *J Biol Chem* 2005, **280**:9595-9603.

69. • Zakharov SD, Hulleman JD, Dutseva EA, Antonenko YN, Rochet JC, Cramer WA: **Helical alpha-synuclein forms highly conductive ion channels**. *Biochemistry* 2007, **46**:14369-14379.
A combined biophysical and electrophysiological study showing that on phospholipid surfaces αSN adopts an α-helical conformation which produces ion channels.

70. •• Tsigelny IF, Bar-On P, Sharikov Y, Crews L, Hashimoto M, Miller MA, Keller SH, Platoshyn O, Yuan JX, Masliah E: **Dynamics of alpha-synuclein aggregation and inhibition of pore-like oligomer development by beta-synuclein**. *FEBS J* 2007, **274**:1862-1877.
A theoretical and experimental study providing evidence that αSN associates to form ring-like α-helical oligomers on the surfaces of phospholipid membranes that produce ion channels, and that the association of αSN with βSN inhibits the process.

71. Hashimoto M, Rockenstein E, Mante M, Crews L, Bar-On P, Gage FH, Marr R, Masliah E: **An antiaggregation gene therapy strategy for Lewy body disease utilizing beta-synuclein lentivirus in a transgenic model**. *Gene Ther* 2004, **11**:1713-1723.