doi:10.1016/j.jmb.2005.04.016 J. Mol. Biol. (2005) 350, 379–392

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Prediction of "Aggregation-prone" and "Aggregation-susceptible" Regions in Proteins Associated with Neurodegenerative Diseases

Amol P. Pawar[1]†, Kateri F. DuBay[1]†, Jesús Zurdo[1], Fabrizio Chiti[2] Michele Vendruscolo[1]* and Christopher M. Dobson[1]*

[1]*Department of Chemistry University of Cambridge Lensfield Road, Cambridge CB2 1EW, UK*

[2]*Dipartimento di Scienze Biochimiche, Viale Morgagni 50 Universitá degli Studi di Firenze 50134 Firenze, Italy*

**Corresponding authors*

Increasing evidence indicates that many peptides and proteins can be converted *in vitro* into highly organised amyloid structures, provided that the appropriate experimental conditions can be found. In this work, we define intrinsic propensities for the aggregation of individual amino acids and develop a method for identifying the regions of the sequence of an unfolded peptide or protein that are most important for promoting amyloid formation. This method is applied to the study of three polypeptides associated with neurodegenerative diseases, Aβ42, α-synuclein and tau. In order to validate the approach, we compare the regions of proteins that are predicted to be most important in driving aggregation, either intrinsically or as the result of mutations, with those determined experimentally. The knowledge of the location and the type of the "sensitive regions" for aggregation is important both for rationalising the effects of sequence changes on the aggregation of polypeptide chains and for the development of targeted strategies to combat diseases associated with amyloid formation.

Keywords: amyloid; sensitive regions; protein aggregation; protein misfolding

Introduction

Biological macromolecules such as proteins, lipids and nucleic acids have the ability to assemble into functional complexes in a highly regulated manner within densely crowded environments.[1,2] Moreover, the balance between normal and pathological self-association has been carefully tuned by molecular evolution.[3,4] Failures of the regulatory mechanisms do, however, occur and may result in conditions such as Alzheimer's and Creutzfeldt–Jakob diseases, and type II diabetes. Such diseases are associated with the deposition in tissue of pathogenic aggregates that are composed largely of misfolded proteins in the form of amyloid fibrils or plaques.[5–10] Recent work suggests that not only are amyloid aggregates formed *in vivo* from a group of otherwise unrelated proteins, but they can be induced *in vitro* from proteins not associated with known deposition diseases.[3,10–13] Such observations have led to the suggestion that the ability to form amyloid fibrils may be a common characteristic of polypeptide chains,[3,5] although individual propensities vary greatly with the sequence and the environmental conditions.

A range of diverse factors has been reported to influence the rate of amyloid formation. Extrinsic factors that can affect the formation of protein aggregates *in vivo* include the interaction with cellular or extracellular components such as molecular chaperones that inhibit misfolding,[14] proteases that frequently generate or process the amyloidogenic precursors,[15] and the effectiveness of quality control mechanisms such as the ubiquitin–proteasome system.[16,17] They also include physicochemical properties that describe the environment of the polypeptide chains, such as pH, temperature, ionic strength, protein concentration, denaturant concentration and pressure.[18–25] Intrinsic factors associated with amyloid formation include a range of fundamental characteristics of polypeptide

† A.P.P. & K.F.D. contributed equally to this work.

Present address: K. F. DuBay, Department of Chemistry, UC Berkeley, 419 Latimer Hall, Berkeley, CA 94720-1460, USA.

E-mail addresses of the corresponding authors: mv245@cam.ac.uk; cmd44@cam.ac.uk

0022-2836/$ - see front matter

chains, such as charge,[26–29] hydrophobicity,[30–32] patterns of polar and non-polar residues,[33] and the propensities to adopt diverse secondary structure elements.[27,32,34,35] In the case of globular proteins, the propensities to form amyloid structures are generally inversely related to the stability of their native states.[24,36–42] Many of the proteins associated with amyloid diseases are, however, at least partially unfolded under physiological conditions. In addition, it is thought that many globular proteins unfold, at least partially, before aggregating. The present study is therefore focused on the conversion of unfolded or partially unfolded states into amyloid aggregates.

One of the most intriguing recent observations in studies of the kinetics of amyloid formation is that polypeptide sequences appear to contain local regions that are "sensitive" for aggregation.[32] Single amino acid mutations in these regions can change the aggregation rates dramatically, while similar changes in other regions may have relatively little effect.[32,43] In addition, it has been shown that it is possible to describe with considerable accuracy the *in vitro* amyloid aggregation propensities of polypeptides using algorithms that take into account the physico-chemical properties of their sequences and of their environment.[44–48] Here, our purpose is to apply this type of analysis to the rationalisation and the prediction of the sensitive regions of polypeptide sequences in general and of proteins associated with neurodegenerative diseases in particular.

Results

Definition of intrinsic aggregation propensities

We define the intrinsic propensity of an unfolded polypeptide chain to form amyloid aggregates, P_{agg}, by considering just the intrinsic factors in the formula that we have recently introduced to define the absolute aggregation rates of unstructured polypeptide chains:[45]

$$P_{agg} = \alpha_{hydr} I^{hydr} + \alpha_{\alpha} I^{\alpha} + \alpha_{\beta} I^{\beta} + \alpha_{pat} I^{pat} + \alpha_{ch} I^{ch} \quad (1)$$

where I^{hydr} represents the hydrophobicity of the sequence,[49,50] I^{α} is the α-helical propensity,[51] I^{β} is the β-sheet propensity,[51] I^{pat} is the hydrophobic patterning,[52] and I^{ch} is the absolute value of the net charge of the sequence; the coefficients α, which weight the individual factors, were determined as described and are listed in Table 1 (see Materials and Methods).[45] Since pH influences some of these terms, such as I^{hydr}, I^{α} and I^{ch}, it must be specified for equation (1), and the hydrophobicity and secondary structure propensities are normalised for the length of the polypeptide chain. The values that we found for the coefficients indicate that hydrophobicity, the presence of specific hydrophobic patterns and the propensity to form β-sheets favour aggregation; in contrast, the net charge and the propensity to form α-helices reduce the tendency to aggregate. It has been suggested that aromatic residues may play an important role in promoting aggregation,[48,53] although other studies have indicated that their importance may be limited.[54] The systematic collection of experimental data relevant to this factor will allow its role to be clarified.

Table 1. Coefficients of equation (1)

Parameter	α	*p*-value
Hydrophobicity	−1.99±0.31	<0.001
α	−5.7±2.3	<0.001
β	5.0±1.7	<0.001
Charge	−0.08±0.03	0.17
Patterns	0.39±0.02	<0.001

The coefficients α used in equation (1). These values and the corresponding statistical errors were derived as described.[45]

Definition of intrinsic *Z*-scores for aggregation

The intrinsic *Z*-score for aggregation, Z_{agg}, enables comparisons to be made between the aggregation propensities of different polypeptide sequences. It is calculated as:

$$Z_{agg} = \frac{P_{agg} - \mu_{agg}}{\sigma_{agg}} \quad (2)$$

where μ_{agg} is the average value of P_{agg} over a set of random polypeptides having the same length as the sequence of interest, and σ_{agg} is the corresponding standard deviation from the average (see Materials and Methods). If $Z_{agg} > 0$, the sequence is more prone to aggregation than a randomly generated sequence at that particular pH, while it is less prone if $Z_{agg} < 0$.

Intrinsic aggregation propensities of individual amino acids

The intrinsic propensities p_{agg} of individual amino acids to promote the conversion of a polypeptide chain into amyloid aggregates can be calculated from equation (1) if one considers a polypeptide sequence of length 1. In this case, the term describing the patterning of polar and non-polar residues, I^{pat}, does not contribute to the resulting p_{agg} value. The aggregation propensity scale obtained in this way should be useful for a qualitative estimate of the effect of mutations on the aggregation behaviour of a given polypeptide chain, at least when hydrophobic patterns are not involved. The scales at three different values of pH are shown in Table 2, where the various amino acids are listed in decreasing order of amyloid formation propensity at neutral pH. At this pH, tryptophan, phenylalanine, cysteine, tyrosine, and isoleucine have the highest amyloid propensities, while

Table 2. Amyloid aggregation propensities of the 20 naturally occurring amino acids

Residue	pH 7	pH 2	pH 13
Trp	2.92	2.92	2.92
Phe	2.80	2.80	2.80
Cys	1.61	1.61	−3.44
Tyr	1.03	1.03	1.03
Ile	0.93	0.93	0.93
Val	0.49	0.49	0.49
Leu	−0.25	−0.25	−0.25
Met	−1.06	−1.06	−1.06
Thr	−2.12	−2.12	−2.12
Ala	−3.31	−3.31	−3.31
Gly	−3.96	−3.96	−3.96
His	−4.31	−9.36	−4.31
Ser	−5.08	−5.08	−5.08
Gln	−6.00	−6.00	−6.00
Asn	−6.02	−6.02	−6.02
Asp	−9.42	−4.38	−9.42
Lys	−9.55	−9.55	−9.47
Glu	−10.38	−6.73	−10.38
Arg	−11.93	−11.93	−11.85
Pro	−11.96	−11.96	−11.96

Amyloid aggregation propensities of individual amino acids, p_{agg}, at three pH values, calculated from equation (1) by considering a polypeptide sequence of length 1.

aspartic acid, lysine, glutamic acid, and arginine have the lowest. The high intrinsic propensities of tryptophan, tyrosine, and phenylalanine are particularly interesting in the light of the studies by Azriel *et al.* that revealed the frequent occurrence of aromatic residues, particularly phenylalanine, in amyloidogenic peptides.[55] The only aromatic residue missing from the amyloidogenic peptides examined is histidine;[55] the scale that we have derived assigns a much lower amyloid aggregation propensity to histidine than to the other aromatic residues, especially at low pH values, where its side-chain is positively charged.

Identification of sensitive regions for aggregation

We use equation (1) to calculate the sum of the individual intrinsic factors (i.e. hydrophobicity, secondary structure propensity, hydrophobic pattern and charge) for each residue in a polypeptide sequence. This operation results in an "aggregation propensity profile", Z_{agg}^{prof} (see Materials and Methods), which illustrates the fact that different local regions in the sequence of a polypeptide chain can have significantly different intrinsic propensities to aggregate. In addition to the profile for the wild-type sequence, at each position we consider all possible amino acid replacements and, by using equation (1), we determine those replacements with the highest and with the lowest aggregation propensity, respectively. In this way, a minimum and a maximum profile can be calculated (see Materials and Methods). From these profiles it is possible to predict the regions where mutations have the greatest potential for increasing or reducing the aggregation propensity. In the

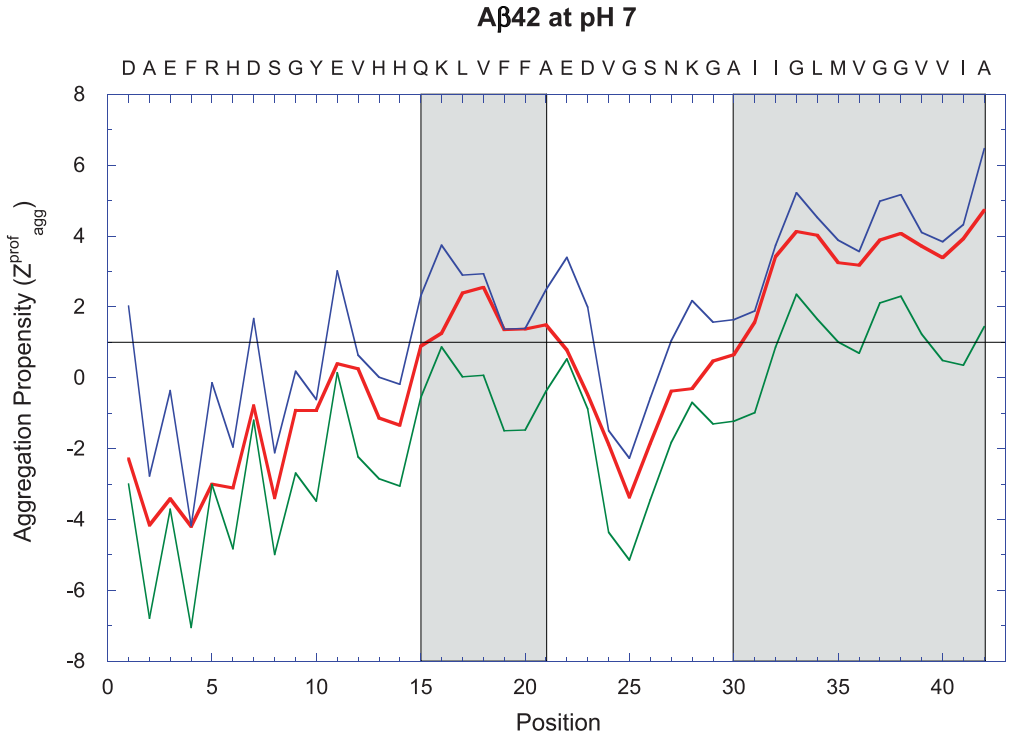

Figure 1. Amyloid aggregation profile of Aβ42 at pH 7. The wild-type profile is plotted in red, along with the maximum (blue) and minimum (green) propensity values at each position along the sequence. A line at $Z_{agg}^{prof} = 1$ is drawn to identify the aggregation-promoting regions; these regions derived experimentally are shown as shaded regions.

following section, we discuss the application of this type of analysis to three polypeptide chains that are natively unfolded and for which there is a considerable amount of experimental data, as their aggregation is associated with Alzheimer's and Parkinson's diseases.

Aβ42

Alzheimer's disease is a neurological disorder characterised by the progressive deposition in the brain of the Aβ40 or the Aβ42 peptide.[9] Since the location, and hence the local pH, at which the aggregation process leading to such deposition occurs *in vivo* is not known precisely, we used equations (1) and (2) to calculate the aggregation propensity profiles for the peptides over a pH range from 2 to 9. Figure 1 shows the result of the calculations at pH 7; the profiles are only weakly dependent on values of pH in the range from 4.5 to 9. By considering all the residues with $Z_{agg}^{prof} > 1.0$, we identified the regions formed by residues 15–21 and 30–42 as those that have the highest local propensities to aggregate.

These predictions are in good agreement with a range of experimental data. Proline scanning mutagenesis has indicated that the region of the sequence spanning residues 15–21 is a particularly important one in the process of amyloid formation.[43] In addition, Tjernberg *et al.* identified several peptides containing residues 16–20 that can readily form fibrils, the shortest of which is 13–23.[56] More recently, Balbach *et al.* have shown that the seven-residue fragment 16–22 is able to form fibrils.[57] The 15–21 region has been defined as part of the structural core of the fibrils, as determined by solid-state NMR and by electron paramagnetic resonance coupled to site-directed spin labelling;[58,59] these two studies indicate that the regions formed by residues 12–24[58] or 13–21[58,59] have a well-defined β-sheet structure in the fibrils. The region of the sequence immediately before the C terminus and spanning residues 31–37 has also been proposed to be important for aggregation, as substitution of residues within this region by proline causes a considerable decrease in aggregation rate.[43] In addition, the peptide corresponding to residues 34–42 can readily form ordered β structures.[60] Solid-state NMR and site-directed spin labelling experiments also suggest that residues 30–40[58] or 30–38[59] form a β-strand in the fibrils. Although the involvement of the last four residues in the cross-β structure of the fibrils is still under debate, there is general agreement that the 31–37 region is part of the core of the fibrillar aggregates.[43,58,59]

We have also used equations (1) and (2) to compute changes in aggregation rate, k, as a consequence of mutation. This approach is of great significance, as many familial amyloid diseases result from mutations within the sequence of the peptide or protein whose aggregation is associated with these medical conditions. At each position of Aβ42, we calculated the average value of $\log (k_{mut}/k_{wt})$ for all 20 different types of amino acids (see Materials and Methods). This type of analysis shows that the regions formed by residues 17–21 and 30–42 have aggregation propensities that cannot be increased significantly by mutation. By contrast, residues 22–23 and 26–29 are the only residues within the 17–42 region of the sequence whose aggregation propensity can be increased significantly by replacement with other residues, as shown in Figure 1. These predictions are in good agreement with the results of a recent study by Wurth *et al.*,[61] in which several mutations were found in the region 29–42 of Aβ42 that can reduce the aggregation propensity significantly. In accord with the experimental findings, we predict that all the single-point variants found by Wurth *et al.* (A2S, L17E, L17Q, F19S, I31N, I32S, I32V, L34P and V32E) have a decreased aggregation propensity (data not shown).[61]

In addition, we examined a series of engineered mutations of Aβ40 and Aβ42 that have been studied *in vitro* by Bitan *et al.*[62] In that study, the aggregation propensities of mutants of Aβ42 at position 41 were ranked as I41G < I41A < I41V < I41L ~ WT; an identical order is predicted by our analysis (Table 3). Additionally, the large oligomers that are known to be formed by the wild-type peptide

Table 3. Amyloid aggregation propensities of Aβ42 mutants

Peptide	Z_{agg}	ΔZ_{agg}	Length[a]
WT	0.61	0.00	42
I41G	0.51	−0.10	42
I41A	0.53	−0.07	42
I41V	0.58	−0.03	42
I41L	0.61	0.00	42
E22Q	0.77	0.17	42
E22G	0.84	0.24	42
E22K	0.78	0.18	42
D23N	0.79	0.18	42
F19P	0.22	−0.39	42
F19T	0.48	−0.13	42
A21G	0.59	−0.02	42
A42G	0.59	−0.02	42
A42V	0.65	0.05	42
Δ (1–2)	0.85	0.25	40
Δ (1–4)	0.97	0.37	38
Δ (1–9)	1.53	0.93	33
D1N	0.79	0.18	42
D7N	0.79	0.18	42
All N	1.12	0.52	42
H6Q	0.57	−0.03	42
H13Q	0.57	−0.03	42
H14Q	0.57	−0.03	42
All Q	0.47	−0.13	42
All N and Q	0.99	0.38	42
15–21	1.61	1.00	7
12–24	1.09	0.48	13
31–37	2.11	1.50	7
30–42	2.52	1.91	13

Amyloid aggregation Z-score propensities of the Aβ42 mutants[62] as determined from equations (1) and (2) at pH 7. ΔZ_{agg} is the change in the Z-score between a mutational variant and the wild-type protein.
[a] Number of amino acid residues.

were not observed in the case of the A42G mutant, while the A42V mutant showed behaviour similar to that of the wild-type peptide.[62] The hypothesis that hydrophobic tendencies drive association of initial aggregates into higher oligomers,[63] is supported by our analysis, which shows a higher Z_{agg} value for the A42V mutant (0.65) relative to the wild-type sequence (0.61) (Table 3). Bitan *et al.* explored the effects of charge neutralisation on the aggregation propensities of Aβ42 by Asp to Asn and by His to Gln mutations.[62] We predict the variant in which all three Asp residues of Aβ42 were mutated (All N) to have a lower aggregation propensity than the variant in which the three Asp and the three His residues were mutated (All N+Q), which in turn is predicted to have a lower aggregation propensity than the wild-type peptide (Table 3). Single Asp to Asn variants (D1N and D7N) are predicted to have a higher aggregation propensity than the All N variant. For the variant in which all the three His residues were mutated (All Q), no significant change in aggregation rate with respect to the wild-type was observed.[62] In this case, we predict a slightly lower aggregation propensity for the His to Gln mutants, because His residues are predominantly uncharged at physiological pH (Table 3).

The aggregation behaviour of the major pathogenic mutants of Aβ40 or Aβ42 that are associated with early-onset Alzheimer's disease or hereditary cerebral haemorrhage with amyloidosis, A21G (Flemish), E22G (Arctic), E22Q (Dutch), E22K (Italian), and D23N (Iowa) have been studied widely *in vitro*.[62,64–66] The Dutch, Arctic, and Iowa mutants all show much higher aggregation rates than Aβ40,[62,64–66] a result that is predicted by our analysis. The Italian mutant was predicted to increase the aggregation propensity, in agreement with the slight increase in the aggregation rate observed experimentally[65] and a higher tendency to oligomerise.[62] The rate of aggregation of the Flemish mutant has been reported in different studies to be decreased slightly or to be similar to that of the wild-type peptide;[66] in addition, the propensity to form oligomers is similar and decreased only slightly when the Flemish mutation is present in Aβ40 and Aβ42, respectively.[62] Our analysis is completely consistent with these observations (Table 3). In addition, the non-pathogenic F19P and F19T mutants have reduced aggregation properties,[62,67,68] in agreement with our predictions (Table 3). Finally, the predicted aggregation propensity for N-terminally truncated peptides, found to occur naturally in amyloid plaques, increases with the extent of truncation, again in accordance with experimental data.[62]

α-Synuclein

α-Synuclein is an intrinsically unstructured protein of 140 residues that is an important component of Lewy bodies, which are intracellular assemblies of aggregates associated with Parkinson's disease and other neurodegenerative disorders known as synucleinopathies.[69] We used equations (1) and (2) to calculate the intrinsic propensity of α-synuclein to aggregate (Table 4). Interestingly, this propensity is significantly lower than that expected for a random polypeptide of the same length. Fragments of α-synuclein are, however, known to be highly amyloidogenic when dissected from the remainder of the sequence, for example 61–95 NAC,[70] 66–74,[71] 69–79[72] and 71–82[73] (Table 4). We predict that all these fragments will indeed have a higher intrinsic propensity to aggregate than a peptide of similar size with a random sequence. Our prediction for the full-length protein is consistent with the expectation that natively unfolded proteins will have evolved to avoid aggregation *in vivo* under normal conditions. Indeed, the amino acid propensity scale for aggregation (Table 2) correlates negatively with the scale that describes the propensity of individual amino acids to appear in unstructured regions.[74] By contrast, the intrinsic aggregation propensities of proteins that are folded in their biologically relevant forms can be much higher, as the regions of high aggregation potential are very likely to be buried in their native states.

The aggregation propensity profile calculated for α-synuclein at pH 7 (Figure 2(a)) reveals that the regions of the sequence with the highest local propensities for aggregation in the wild-type protein are 38–40, 50–54, 65–75 and 87–92. All these regions are in a large fragment of the protein (residues 34–101) that has been shown by site-directed spin-labelling to form the structural core of the fibril.[75] We predict that this fragment will have a higher propensity to aggregate than a random sequence of similar length ($Z_{agg}=0.29$), and a much higher propensity than the intact protein ($Z_{agg}=-1.68$, Table 4). Further, two independent studies have indicated that a stretch of the sequence in the central region of the sequence encompassing residues 69–79[72] (or 71–82)[73] is a key region for promoting aggregation. This part of the sequence is located within the proposed fibril core and corresponds to part of the most prominent peak in our

Table 4. Amyloid aggregation propensities of α-synuclein

Sequence	Z_{agg}	Length[a]
Wild-type	−1.68	140
A30P	−1.75	140
A53T	−1.67	140
E46K	−1.53	140
NAC (61–95)	1.42	35
69–79	1.81	11
66–74	1.89	9
34–101 (fibril core)	0.29	68
71–82	1.46	12
36–101	0.51	66
NACP112	−0.46	112

Amyloid aggregation Z-score propensities of α-synuclein and several of its mutants and fragments, determined from equations (1) and (2) at pH 7.
[a] Number of amino acid residues.

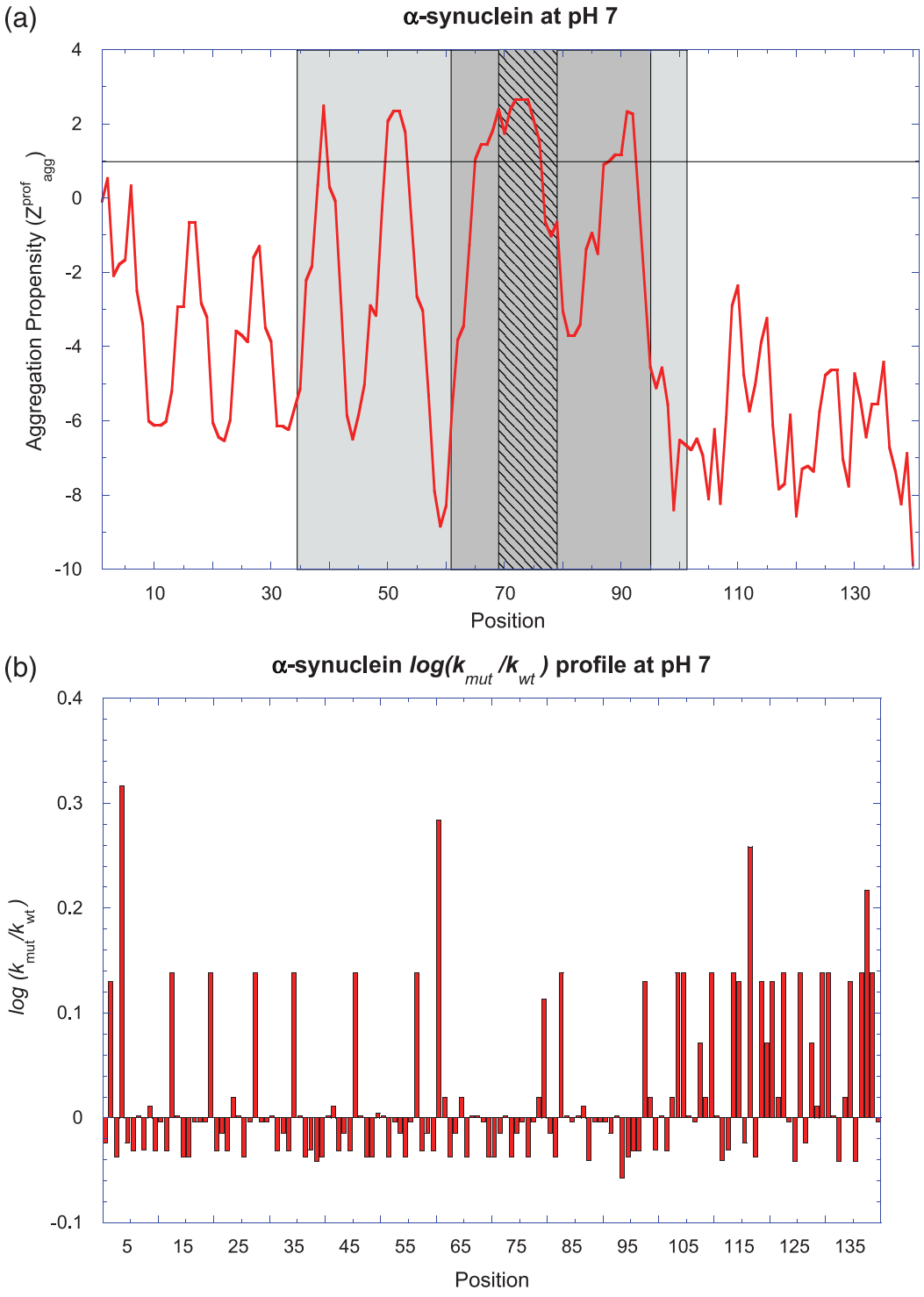

Figure 2. (a) Amyloid aggregation profile of α-synuclein at pH 7. A line at $Z^{prof}_{agg} = 1$ is drawn to identify the sensitive regions. The large region of the protein thought to be structured in the fibrils is shown by pale grey.[75] The highly amyloidogenic NAC region[70] is shown by dark grey and the 69–79 region, found to be a particularly amyloidogenic segment within the NAC region,[72] is shown by hashed lines. (b) Plot of the log (k_{mut}/k_{wt}) profile (see Materials and Methods) at pH 7 for the 20 possible amino acid types at each position of the sequence of α-synuclein.

aggregation propensity profile (Figure 2(a)). One might imagine that the 69–79 (or 71–82) region, perhaps in conjunction with other short segments of the sequence, would initiate the process of aggregation that subsequently extends to include much of the rest of the protein molecule.

The log (k_{mut}/k_{wt}) profiles (see Materials and Methods) for the 20 possible amino acid types at each position in the sequence were calculated for α-synuclein (Figure 2(b)). A striking feature is the presence of peaks associated with Lys and Arg residues that are almost evenly spaced throughout

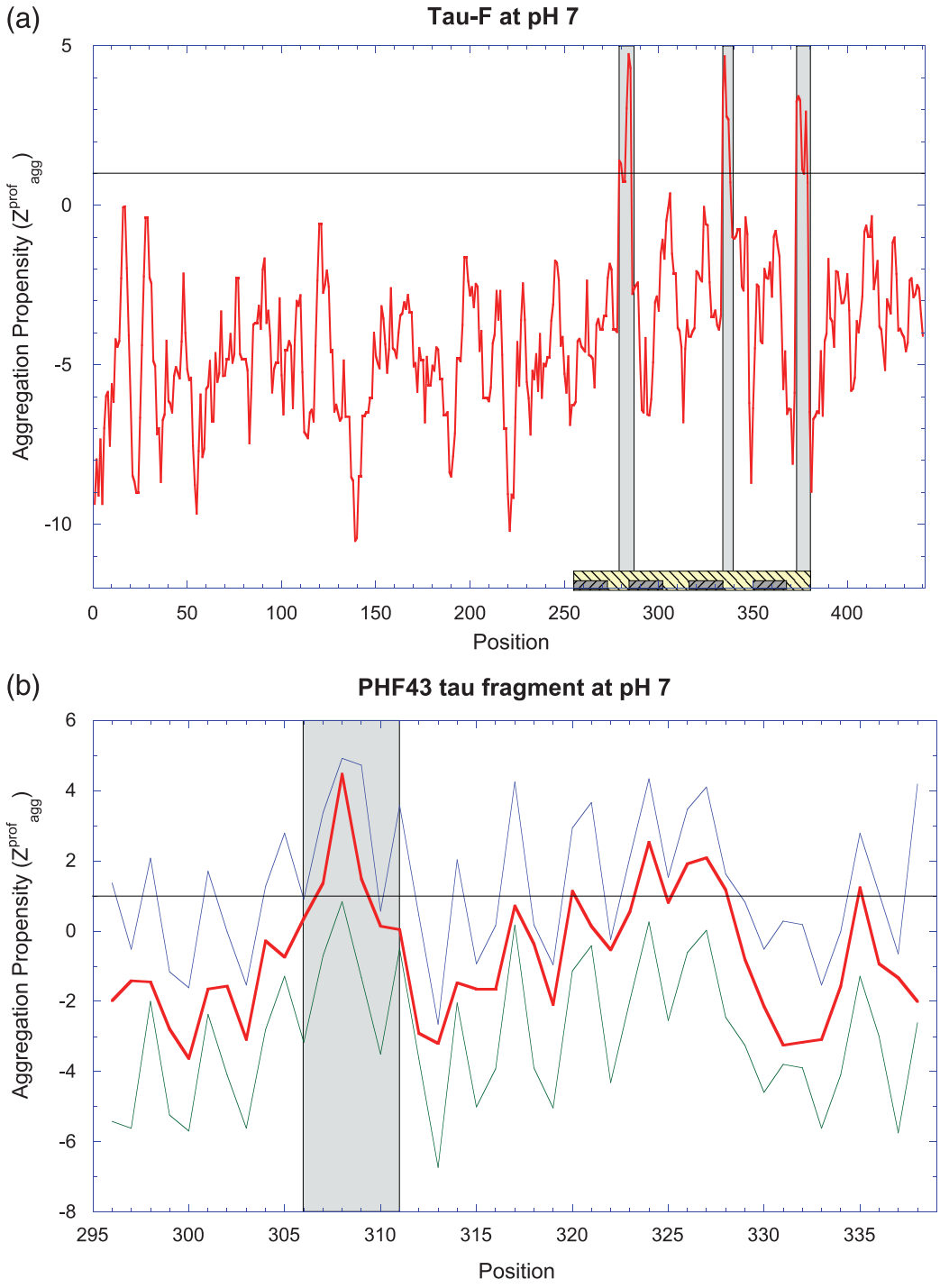

Figure 3. (a) Aggregation propensity profile of tau at pH 7. (b) Aggregation propensity profile of the PHF43 fragment at pH 7. The wild-type profile is plotted in red, along with the maximum (blue) and minimum (green) possible propensity values for each residue. In (a) the microtubule-binding region and the four microtubule-binding repeats are indicated by hashed areas and the sensitive regions that we identified are indicated as shaded areas. In (b) the aggregation promoting region derived experimentally is shown as shaded regions and a line at $Z^{prof}_{agg} = 1$ is drawn to identify the aggregation-promoting regions.

the sequence. We suggest that they play the role of "sequence breakers"[32,76–78] to reduce the overall aggregation propensity of the protein by inhibiting self-association of regions of the sequence that contain such charged residues. In support of this suggestion, the NAC region is the only part of the sequence where such sequence breakers are absent. Sequence breakers may be important in disrupting hydrophobic regions of the sequence; for example, the substitution of a residue that disrupts a hydrophobic pattern will decrease significantly the aggregation propensity of a polypeptide sequence.

Two familial variants of α-synuclein, A30P[79] and

Table 5. Amyloid aggregation propensities of tau mutants

Peptide	Z_{agg}	ΔZ_{agg}
WT	−2.23	0.00
R5L	−2.15	0.08
R5H	−2.16	0.06
K257T	−2.17	0.05
L266V	−2.22	0.00
G272V	−2.22	0.01
N279K	−2.27	−0.05
N296H	−2.22	0.01
P301L	−2.22	0.00
P301S	−2.21	0.02
S305N	−2.23	0.00
S320F	−1.98	0.24
V337M	−2.23	0.00
E342V	−1.80	0.42
K369I	−1.93	0.29
G389R	−2.29	−0.07
R406W	−2.15	0.08
ΔK280	−2.16	0.06
ΔN296	−2.22	0.01

Amyloid aggregation Z-score propensities of tau mutants[90,95] as determined from equations (1) and (2) at pH 7. ΔZ_{agg} is the change in the Z-score between a mutant and the wild-type protein.

A53T,[80] have been found to be associated with the early onset of Parkinson's disease. These substitutions are not, however, predicted to perturb the aggregation rates significantly, as the substituted residues generate intrinsic aggregation propensities similar to that of the wild-type protein (Table 4). These predictions can be compared with *in vitro* experimental studies that show that A30P has an aggregation rate similar to that of the wild-type protein, while A53T forms oligomers and fibrils slightly more rapidly.[81] These findings suggest that the pathogenic effects of these two mutations could arise from factors other than changes in intrinsic aggregation properties. Indeed, the pathogenicity of the A30P and A53T mutants has been suggested to be due to the disruption of the transport of α-synuclein, with the rate of transport being much slower than that of the wild-type protein.[82] The slower transport would lead to the creation of high local concentrations of protein that will undoubtedly promote a faster protofibril formation. This conclusion suggests that the ability to predict aggregation propensities using an approach such as that discussed here could be of great value in identifying factors that can reduce or enhance aggregation of a protein by mechanisms other than those that perturb its intrinsic rate of aggregation. E46K is the latest mutation in α-synuclein that has been linked to Parkinson's disease.[83] Biophysical characterisation of this mutant showed an increased aggregation propensity relative to wild-type α-synuclein.[84] In agreement with the experimental observations, our calculations predict an enhanced aggregation propensity for the mutant compared to that of the wild-type protein.

Alternative spliced forms of α-synuclein mRNA have been described.[85] Although their role in Parkinson's disease is not clear, recent evidence has suggested that differences in the expression levels of α-synuclein variants could have an impact in some pathologies associated with the presence of Lewy bodies.[86] Our predictions suggest that the alternatively spliced form α-synuclein NACP112 should have a very considerably increased aggregation propensity. These results are in agreement with recent structural studies that suggested that the C terminus region of the full-length α-synuclein, which is known to inhibit its aggregation,[87,88] has the tendency to interact with the NAC region, thus partially shielding it from forming intermolecular aggregates.[89]

tau

The protein tau, a microtubule-associated polypeptide representing an important component of the neuronal cytoskeleton, is natively unfolded in its full-length form.[90,91] Tau has been identified as a major component of the intracellular neurofibrillary tangles associated with Alzheimer's disease,[92] and has thus attracted considerable attention.[90,91] Tau-mediated cytotoxicity may arise from overexpression,[93] hyperphosphorylation[94] or proteolysis.[94] The aggregation propensity profile for the longest isoform of tau (tau-F, 441 amino acid residues), calculated using equations (1) and (2), is shown in Figure 3(a). The plot shows that three local regions of the sequence are predicted to be highly prone to aggregation; these are 280–287, 335–340 and 374–381. These regions are in the microtubule-binding region, as indicated in Figure 3(a). These regions are located in the segment of tau (residues 257–389) where most point mutants linked to FTDP-17 are found.

A recent review by Gamblin *et al.* lists 17 pathogenic missense point mutants in tau,[90] and an additional one is described by Goedert *et al.*[95] The intrinsic propensities of wild-type and mutant forms of tau have been calculated using equations (1) and (2), and are reported in Table 5. Twelve of these mutants are predicted to have a Z_{agg} score higher than that of wild-type tau, indicating that the mutations have increased the intrinsic rate of aggregation. Two of the 18 mutations, N279K and G389R, are predicted to have a lower score, but they are both known to cause changes to factors other than those affecting the aggregation rate. In particular, N279K enhances the splicing of exon 10,[95] and thereby increases the expression of the isoforms containing four repeats in the microtubule-binding domain. These isoforms have a decreased ability to bind microtubules, leading to a higher local concentration of free tau, a factor that is expected to increase the aggregation rate by itself.[95] The G389R substitution also decreases the ability of tau to bind to microtubules,[90] suggesting that similar factors give rise to its pathogenicity. The other four mutants, S305N, L266V, P301L and V337M, are predicted to have approximately the same aggregation propensities as the wild-type protein. Of these, S305N has an effect similar to N279K, and V337M is

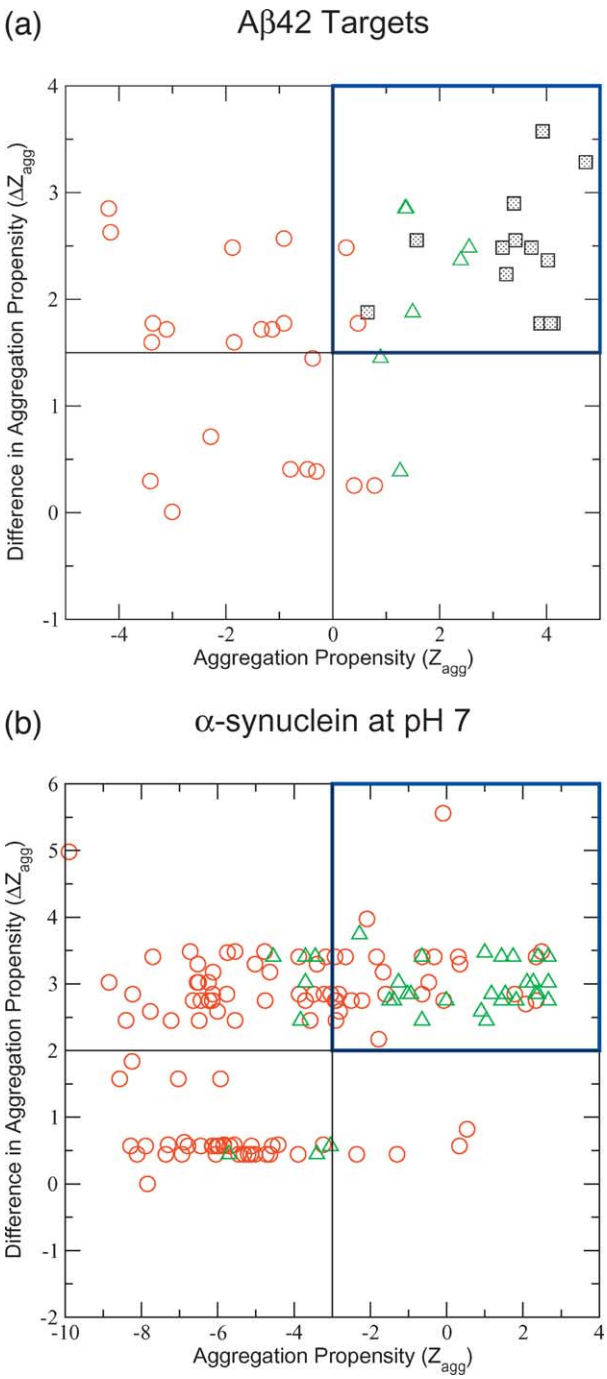

Figure 4. Type $S-$ regions are segments of the polypeptide chain that are predicted to have a high propensity to aggregate but that can become of low propensity upon mutation. We illustrate this behaviour by showing, for Aβ42 and α-synuclein, for each position along the sequence, the maximum predicted decrease of the aggregation propensity upon mutation (y-axis) against the propensity for aggregation of the wild-type sequence (x-axis). Residues in the upper right quadrant have a high aggregation propensity but can change to being of low propensity upon mutation. (a) Aβ42; residues in the regions 15–21 (triangles) and 30–42 (squares) cluster in the upper right quadrant, showing that they form a type $S-$ region. (b) α-Synuclein; residues in the NAC fragment also form a type $S-$ region (triangles), and also cluster in the upper right quadrant.

the only FTDP-17 mutant that reduces the β-sheet propensity;[90] the origins of pathogenicity are not known for the remaining three mutations. Nevertheless, these results suggest that the pathogenicity of at least 12 of the 18 familial mutations could arise, at least in part, from an increase in their rates of aggregation relative to the wild-type protein.

A recent report has suggested that the minimal interaction motif of tau that gives rise to aggregation is a 43 residue segment (PHF43, or N265-E338ΔR2) corresponding to the third microtubule-binding repeat sequence along with a number of adjacent residues.[96] In addition, a spot membrane-binding assay was used to suggest that a six residue segment within PHF43 (PHF6, or ^{306}VQIVYK^{311}) is the specific region of this peptide that promotes aggregation.[96] Calculation of the aggregation propensity profile of PHF43 indicates that the highest peak in the profile, identified by using a threshold $Z_{agg}^{prof} = 1$, does indeed correspond to the PHF6 segment (Figure 3(b)).

Discussion and Conclusions

Here, we describe a method for calculating the intrinsic amyloid aggregation propensities of polypeptide sequences, and we have used this approach to calculate the aggregation propensity profiles for three natively unfolded polypeptide chains associated with neurodegenerative diseases, Aβ42, α-synuclein and tau. These profiles have allowed the regions of these sequences that influence their aggregation behaviour to be identified and compared to the results of experimental studies.

We can rationalise the results by identifying two types of "sensitive" regions. We define the first type as aggregation-prone, or P regions. These are local regions within a sequence that are predicted to have a high intrinsic aggregation propensity and hence are likely to promote the aggregation of the entire sequence. In addition, they are likely to be part of the β-sheet core in the resulting amyloid protofibrils and fibrils. It is important to identify such regions, as they may represent very favourable targets for therapeutic strategies based on their shielding, for example through competitive binding.[97,98] We define the second type of sensitive regions as aggregation-susceptible, or S regions in which large changes in the propensities for aggregation are predicted to be possible as a result of mutations. These regions can be identified from the analysis of the minimum and the maximum propensity profiles, such as those shown in Figures 1 and 3(b). More specifically, we can expect two types of S regions, depending on their response to single amino acid mutations: regions of low propensity that can become of high propensity (type $S+$) and regions of high propensity that can become of low propensity (type $S-$). Regions of type $S+$ are likely to be important in familial forms of misfolding diseases, as mutations localised in these regions can increase the propensity for aggregation

substantially. Residues 22–23 in Aβ42 and 100–106 in α-synuclein are prominent examples of type $S+$ regions. Regions of type $S-$ are suitable targets for future strategies for countering aggregation based on gene therapy or stem-cell implantation,[34] as mutations in these regions are expected to be potentially highly favourable for reducing the aggregation propensity.[99] Residues 30–42 of Aβ42, the NAC region of α-synuclein and the 306–311 region of tau are prominent examples of regions of type $S-$ (Figure 4).

Overall, therefore, we have shown that it is possible to use arguments based on physico-chemical principles to identify the regions of the sequence of an unstructured peptide or natively unfolded protein that are most important for promoting aggregation. The current approach is designed to be directly relevant for unstructured polypeptide chains, such as those associated with neurological disorders, including Alzheimer's and Parkinson's diseases, and non-neuropathic disorders, such as type II diabetes. An additional term representing the stability of the native state can in principle be included to extend the approach to proteins that aggregate from a predominantly globular state, as is the case, for example, in the systemic amyloidoses associated with lysozyme and transthyretin.[10] Moreover, and more generally, it should be possible to include in the predictive algorithm other factors that will emerge as important determinants of aggregation propensity of any given system.

Knowledge of the sensitive regions for amyloid aggregation, and the ability to predict their response to various types of perturbations, including amino acid replacements and changes in pH, are of particular significance for identifying potential disease-associated mutations and for the development of rational approaches to engineering sequences with altered aggregation propensities. Such knowledge will be crucial in many approaches to combat protein deposition diseases, in particular those involving the identification of small molecules[98] or antibodies[97,100] that are designed to inhibit the formation of pathogenic aggregates.

Materials and Methods

Fitting of the coefficients using known aggregation rates

The intrinsic factors included in the algorithm developed[45] were used in this work to calculate P_{agg}, the intrinsic aggregation propensity of a sequence. The weights of the various intrinsic factors were determined simultaneously by using regression techniques using an extended database of 203 sequences that included the 83 sequences used by DuBay *et al.*[45] We used normalised β-sheet and α-helix propensity scales,[51] with the following modifications: for β-sheet propensity calculations, we set Pro = 0.1, to account for the known ability of this residue to disrupt β-sheets; for α-helix propensity calculations, we set Gly = 2 and Pro = 1.5, to avoid zero values in the logarithm.

Amino acid aggregation propensities

We calculated the propensity for aggregation, p_{agg}, for individual amino acids (Table 2) by applying equation (1) to a sequence comprising just one amino acid residue; in this case the term I^{pat} is always equal to zero.

Determination of the aggregation propensities of random peptides

Using the SWISS-PROT scale of the relative occurrence of different amino acids in polypeptide sequences, the intrinsic aggregation propensity of a random polypeptide of a given length was determined by averaging P_{agg} over 100,000 randomly generated sequences of that length.

Determination of the aggregation propensity profiles

We used equation (1) to calculate the p_{agg} values of individual amino acids at each position of a polypeptide sequence in order to define the propensities of local regions of polypeptide chains. In this case, we set $I^{pat} = 1$ for all residues within a five residue sequence pattern of alternating hydrophobic and hydrophilic residues. To facilitate the comparison of different sequences, we plotted $Z_{agg}^{prof} = (p_{agg} - \mu_{agg})/\sigma_{agg}$, where μ_{agg} is the average P_{agg} value of a random polypeptide chain of the same length and σ_{agg} is the corresponding standard deviation. The resulting Z_{agg}^{prof} profile was then smoothed by averaging over a sliding window of seven residues.

Determination of the maximum and minimum aggregation propensity profiles

The regions of the sequence for which the amyloid aggregation rates of a polypeptide chain are particularly liable to change as a result of single amino acid mutations were identified as follows. At each position in the sequence, the Z_{agg}^{prof} values were calculated for each of the 19 possible changes of the type of amino acid relative to the wild-type sequence. Each of these 19 possibilities corresponds to a single-residue mutation of the wild-type sequence. The maximum and the minimum aggregation profiles were then constructed by considering the highest and the lowest values of the aggregation propensities for the 19 mutations at each position. These two profiles can thus be compared with the profile for the wild-type sequence to identify the range of aggregation rates resulting from substitutions at each position in the sequence.

Determination of the log (k_{mut}/k_{wt}) profiles

At each position in a polypeptide sequence, we considered the wild-type residue and the 19 possible single-residue mutants and calculated the differences in the logarithm of the aggregation rates between the mutants and the wild-type sequences by using the formula:

$$\log(k_{mut}/k_{wt}) = P_{agg}(\text{mut}) - P_{agg}(\text{wt})$$

Peaks in the log (k_{mut}/k_{wt}) profiles indicate regions of the sequence where mutations have a high probability of increasing the rate of aggregation.

Acknowledgements

We are grateful for support from the Gates Cambridge Trust (A.P.P., K.F.D.), the Leverhulme Trust (to J.Z., M.V. and C.M.D.), the Italian MIUR and CNR (to F.C.), the Royal Society (to M.V.) and the Wellcome Trust (to C.M.D.).

References

1. Ellis, R. J. (2001). Macromolecular crowding: obvious but underappreciated. *Trends Biochem.Sci.* **26**, 597–604.
2. Minton, A. P. (2000). Implications of macromolecular crowding for protein assembly. *Curr. Opin. Struct. Biol.* **10**, 34–39.
3. Dobson, C. M. (1999). Protein misfolding, evolution and disease. *Trends Biochem. Sci.* **24**, 329–332.
4. Vendruscolo, M., Zurdo, J., MacPhee, C. E. & Dobson, C. M. (2003). Protein folding and misfolding: a paradigm of self-assembly and regulation in complex biological systems. *Phil. Trans. Roy. Soc. ser. A*, **361**, 1205–1222.
5. Dobson, C. M. (2001). The structural basis of protein folding and its links with human disease. *Phil. Trans. Roy. Soc. ser. B*, **356**, 133–145.
6. Horwich, A. L. & Weissman, J. S. (1997). Deadly conformations—protein misfolding in prion disease. *Cell*, **89**, 499–510.
7. Kelly, J. W. (1998). The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. *Curr. Opin. Struct. Biol.* **8**, 101–106.
8. Rochet, J. C. & Lansbury, P. T. (2000). Amyloid fibrillogenesis: themes and variations. *Curr. Opin. Struct. Biol.* **10**, 60–68.
9. Selkoe, D. J. (2002). Alzheimer's disease is a synaptic failure. *Science*, **298**, 789–791.
10. Stefani, M. & Dobson, C. M. (2003). Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *J. Mol. Med.* **81**, 678–699.
11. Chiti, F., Webster, P., Taddei, N., Clark, A., Stefani, M., Ramponi, G. & Dobson, C. M. (1999). Designing conditions for *in vitro* formation of amyloid protofilaments and fibrils. *Proc. Natl Acad. Sci. USA*, **96**, 3590–3594.
12. Guijarro, J. I., Sunde, M., Jones, J. A., Campbell, I. D. & Dobson, C. M. (1998). Amyloid fibril formation by an SH3 domain. *Proc. Natl Acad. Sci. USA*, **95**, 4224–4228.
13. Uversky, V. N. & Fink, A. L. (2004). Conformational constraints for amyloid fibrillation: the importance of being unfolded. *BBA-Proteins Proteomics*, **1698**, 131–153.
14. Muchowski, P. J. (2002). Protein misfolding, amyloid formation, and neurodegeneration: a critical role for molecular chaperones? *Neuron*, **35**, 9–12.
15. Citron, M., Westaway, D., Xia, W. M., Carlson, G., Diehl, T., Levesque, G. *et al*. (1997). Mutant presenilins of Alzheimer's disease increase production of 42-residue amyloid beta-protein in both transfected cells and transgenic mice. *Nature Med.* **3**, 67–72.
16. Bence, N. F., Sampat, R. M. & Kopito, R. R. (2001). Impairment of the ubiquitin-proteasome system by protein aggregation. *Science*, **292**, 1552–1555.
17. Tofaris, G. K., Razzaq, A., Ghetti, B., Lilley, K., Spillantini, M. G. (2003). Ubiquitination of alpha-synuclein in Lewy bodies is a pathological event not associated with impairment of proteasome function. *J. Biol. Chem.* **298**, 44405–44411.
18. Abrahamson, M. & Grubb, A. (1994). Increased body-temperature accelerates aggregation of the Leu-68-Gln mutant cystatin-C, the amyloid-forming protein in hereditary cystatin-C amyloid angiopathy. *Proc. Natl Acad. Sci. USA*, **91**, 1416–1420.
19. Lomakin, A., Chung, D. S., Benedek, G. B., Kirschner, D. A. & Teplow, D. B. (1996). On the nucleation and growth of amyloid beta-protein fibrils: detection of nuclei and quantitation of rate constants. *Proc. Natl Acad. Sci. USA*, **93**, 1125–1129.
20. Harper, J. D. & Lansbury, P. T. (1997). Models of amyloid seeding in Alzheimier's disease and scrapie: mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins. *Annu. Rev. Biochem.* **66**, 385–407.
21. Kusumoto, Y., Lomakin, A., Teplow, D. B. & Benedek, G. B. (1998). Temperature dependence of amyloid beta-protein fibrillization. *Proc. Natl Acad. Sci. USA*, **95**, 12277–12282.
22. Su, Y. & Chang, P. T. (2001). Acidic pH promotes the formation of toxic fibrils from beta-amyloid peptide. *Brain Res.* **893**, 287–291.
23. Zurdo, J., Guijarro, J. I., Jimenez, J. L., Saibil, H. R. & Dobson, C. M. (2001). Dependence on solution conditions of aggregation and amyloid formation by an SH3 domain. *J. Mol. Biol.* **311**, 325–340.
24. Chiti, F., Taddei, N., Bucciantini, M., White, P., Ramponi, G. & Dobson, C. M. (2000). Mutational analysis of the propensity for amyloid formation by a globular protein. *EMBO J.* **19**, 1441–1449.
25. Ferrao-Gonzales, A. D., Souto, S. O., Silva, J. L. & Foguel, D. (2000). The preaggregated state of an amyloidogenic protein: hydrostatic pressure converts native transthyretin into the amyloidogenic state. *Proc. Natl Acad. Sci. USA*, **97**, 6445–6450.
26. Konno, T. (2001). Amyloid-induced aggregation and precipitation of soluble proteins: a electrostatic contribution of the Alzheimer's beta(25–35) amyloid fibril. *Biochemistry*, **40**, 2148–2154.
27. Tjernberg, L., Hosia, W., Bark, N., Thyberg, J. & Johansson, J. (2002). Charge attraction and beta propensity are necessary for amyloid fibril formation from tetrapeptides. *J. Biol. Chem.* **277**, 43243–43246.
28. Chiti, F., Calamai, M., Taddei, N., Stefani, M., Ramponi, G. & Dobson, C. M. (2002). Studies of the aggregation of mutant proteins *in vitro* provide insights into the genetics of amyloid diseases. *Proc. Natl Acad. Sci. USA*, **99**, 16419–16426.
29. de la Paz, M. L., Goldie, K., Zurdo, J., Lacroix, E., Dobson, C. M., Hoenger, A. & Serrano, L. (2002). *De novo* designed peptide-based amyloid fibrils. *Proc. Natl Acad. Sci. USA*, **99**, 16052–16057.
30. Otzen, D. E., Kristensen, O. & Oliveberg, M. (2000). Designed protein tetramer zipped together with a hydrophobic Alzheimer homology: a structural clue to amyloid assembly. *Proc. Natl Acad. Sci. USA*, **97**, 9907–9912.
31. Schwartz, R., Istrail, S. & King, J. (2001). Frequencies of amino acid strings in globular protein sequences indicate suppression of blocks of consecutive hydrophobic residues. *Protein Sci.* **10**, 1023–1031.
32. Chiti, F., Taddei, N., Baroni, F., Capanni, C., Stefani,

M., Ramponi, G. & Dobson, C. M. (2002). Kinetic partitioning of protein folding and aggregation. *Nature Struct. Biol.* **9**, 137–143.
33. West, M. W., Wang, W. X., Patterson, J., Mancias, J. D., Beasley, J. R. & Hecht, M. H. (1999). *De novo* amyloid proteins from designed combinatorial libraries. *Proc. Natl Acad. Sci. USA*, **96**, 11211–11216.
34. Villegas, V., Zurdo, J., Filimonov, V. V., Aviles, F. X., Dobson, C. M. & Serrano, L. (2000). Protein engineering as a strategy to avoid formation of amyloid fibrils. *Protein Sci.* **9**, 1700–1708.
35. Kallberg, Y., Gustafsson, M., Persson, B., Thyberg, J. & Johansson, J. (2001). Prediction of amyloid fibril-forming proteins. *J. Biol. Chem.* **276**, 12945–12950.
36. Hurle, M. R., Helms, L. R., Li, L., Chan, W. N. & Wetzel, R. (1994). A role for destabilizing amino-acid replacements in light-chain amyloidosis. *Proc. Natl Acad. Sci. USA*, **91**, 5446–5450.
37. Quintas, A., Saraiva, M. J. M. & Brito, R. M. M. (1999). The tetrameric protein transthyretin dissociates to a non-native monomer in solution—a novel model for amyloidogenesis. *J. Biol. Chem.* **274**, 32943–32949.
38. Ramirez-Alvarado, M., Merkel, J. S. & Regan, L. (2000). A systematic exploration of the influence of the protein stability on amyloid fibril formation *in vitro*. *Proc. Natl Acad. Sci. USA*, **97**, 8979–8984.
39. Siepen, J. A. & Westhead, D. R. (2002). The fibril_one on-line database: mutations, experimental conditions, and trends associated with amyloid fibril formation. *Protein Sci.* **11**, 1862–1866.
40. Smith, D. P., Jones, S., Serpell, L. C., Sunde, M. & Radford, S. E. (2003). A systematic investigation into the effect of protein destabilisation on beta 2-microglobulin amyloid formation. *J. Mol. Biol.* **330**, 943–954.
41. Booth, D. R., Sunde, M., Bellotti, V., Robinson, C. V., Hutchinson, W. L., Fraser, P. E. *et al*. (1997). Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature*, **385**, 787–793.
42. Stathopulos, P. B., Rumfeldt, J. A. O., Scholz, G. A., Irani, R. A., Frey, H. E., Hallewell, R. A. *et al*. (2003). Cu/Zn superoxide dismutase mutants associated with amyotrophic lateral sclerosis show enhanced formation of aggregates *in vitro*. *Proc. Natl Acad. Sci. USA*, **100**, 7021–7026.
43. Williams, A. D., Portelius, E., Kheterpal, I., Guo, J. T., Cook, K. D., Xu, Y. & Wetzel, R. (2004). Mapping A beta amyloid fibril secondary structure using scanning proline mutagenesis. *J. Mol. Biol.* **335**, 833–842.
44. Chiti, F., Stefani, M., Taddei, N., Ramponi, G. & Dobson, C. M. (2003). Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature*, **424**, 805–808.
45. Dubay, K. F., Pawar, A. P., Chiti, F., Zurdo, J., Dobson, C. M. & Vendruscolo, M. (2004). Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains. *J. Mol. Biol.* **341**, 1317–1326.
46. Fernandez-Escamilla, A. M., Rousseau, F., Schymkowitz, J. & Serrano, L. (2004). Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nature Biotechnol.* **22**, 1302–1306.
47. Linding, R., Schymkowitz, J., Rousseau, F., Diella, F. & Serrano, L. (2004). A comparative study of the relationship between protein structure and beta-aggregation in globular and intrinsically disordered proteins. *J. Mol. Biol.* **342**, 345–353.
48. Tartaglia, G. G., Cavalli, A., Pellarin, R. & Caflisch, A. (2004). The role of aromaticity, exposed surface, and dipole moment in determining protein aggregation rates. *Protein Sci.* **13**, 1939–1941.
49. Cowan, R. & Witthaker, R. G. (1990). Hydrophobicity indices for amino acid residues as determined by HPLC. *Pept. Res.* **3**, 75–80.
50. Roseman, M. A. (1988). Hydrophilicity of polar amino-acid side-chains is markedly reduced by flanking peptide-bonds. *J. Mol. Biol.* **200**, 513–522.
51. Koehl, P. & Levitt, M. (1999). Structure-based conformational preferences of amino acids. *Proc. Natl Acad. Sci. USA*, **96**, 12524–12529.
52. Broome, B. M. & Hecht, M. H. (2000). Nature disfavors sequences of alternating polar and non-polar amino acids: implications for amyloidogenesis. *J. Mol. Biol.* **296**, 961–968.
53. Makin, O. S., Atkins, E., Sikorski, P., Johansson, J. & Serpell, L. C. (2005). Molecular basis for amyloid fibril formation and stability. *Proc. Natl Acad. Sci. USA*, **102**, 315–320.
54. Tracz, S. M., Abedini, A., Driscoll, M. & Raleigh, D. P. (2004). Role of aromatic interactions in amyloid formation by peptides derived from human amylin. *Biochemistry*, **43**, 15901–15908.
55. Azriel, R. & Gazit, E. (2001). Analysis of the minimal amyloid-forming fragment of the islet amyloid polypeptide—an experimental support for the key role of the phenylalanine residue in amyloid formation. *J. Biol. Chem.* **276**, 34156–34161.
56. Tjernberg, L. O., Callaway, D. J. E., Tjernberg, A., Hahne, S., Lilliehook, C., Terenius, L. *et al*. (1999). A molecular model of Alzheimer amyloid beta-peptide fibril formation. *J. Biol. Chem.* **274**, 12619–12625.
57. Balbach, J. J., Ishii, Y., Antzutkin, O. N., Leapman, R. D., Rizzo, N. W., Dyda, F. *et al*. (2000). Amyloid fibril formation by A beta(16–22), a seven-residue fragment of the Alzheimer's beta-amyloid peptide, and structural characterization by solid state NMR. *Biochemistry*, **39**, 13748–13759.
58. Petkova, A. T., Ishii, Y., Balbach, J. J., Antzutkin, O. N., Leapman, R. D., Delaglio, F. & Tycko, R. (2002). A structural model for Alzheimer's beta-amyloid fibrils based on experimental constraints from solid state NMR. *Proc. Natl Acad. Sci. USA*, **99**, 16742–16747.
59. Torok, M., Milton, S., Kayed, R., Wu, P., McIntire, T., Glabe, C. G. & Langen, R. (2002). Structural and dynamic features of Alzheimer's A beta peptide in amyloid fibrils studied by site-directed spin labeling. *J. Biol. Chem.* **277**, 40810–40815.
60. Lansbury, P. T., Costa, P. R., Griffiths, J. M., Simon, E. J., Auger, M., Halverson, K. J. *et al*. (1995). Structural model for the beta-amyloid fibril based on interstrand alignment of an antiparallel-sheet comprising a C-terminal peptide. *Nature Struct. Biol.* **2**, 990–998.
61. Wurth, C., Guimard, N. K. & Hecht, M. H. (2002). Mutations that reduce aggregation of the Alzheimer's A beta 42 peptide: an unbiased search for the sequence determinants of A beta amyloidogenesis. *J. Mol. Biol.* **319**, 1279–1290.
62. Bitan, G., Vollers, S. S. & Teplow, D. B. (2003).

Elucidation of primary structure elements controlling early amyloid beta-protein oligomerization. *J. Biol. Chem.* **278**, 34882–34889.
63. Bitan, G., Lomakin, A. & Teplow, D. B. (2001). Amyloid beta-protein oligomerization—prenucleation interactions revealed by photo-induced cross-linking of unmodified proteins. *J. Biol. Chem.* **276**, 35176–35184.
64. Nilsberth, C., Westlind-Danielsson, A., Eckman, C. B., Condron, M. M., Axelman, K., Forsell, C. *et al*. (2001). The "Arctic" APP mutation (E693G) causes Alzheimer's disease by enhanced A beta protofibril formation. *Nature Neurosci.* **4**, 887–893.
65. Miravalle, L., Tokuda, T., Chiarle, R., Giaccone, G., Bugiani, O., Tagliavini, F. *et al*. (2000). Substitutions at codon 22 of Alzheimer's A beta peptide induce diverse conformational changes and apoptotic effects in human cerebral endothelial cells. *J. Biol. Chem.* **275**, 27110–27116.
66. Van Nostrand, W. E., Melchor, J. P., Cho, H. S., Greenberg, S. M. & Rebeck, G. W. (2001). Pathogenic effects of D23N Iowa mutant amyloid beta-protein. *J. Biol. Chem.* **276**, 32860–32866.
67. Esler, W. P., Stimson, E. R., Ghilardi, J. R., Lu, Y. A., Felix, A. M., Vinters, H. V. *et al*. (1996). Point substitution in the central hydrophobic cluster of a human beta-amyloid congener disrupts peptide folding and abolishes plaque competence. *Biochemistry*, **35**, 13914–13921.
68. Wood, S. J., Wetzel, R., Martin, J. D. & Hurle, M. R. (1995). Prolines and amyloidogenicity in fragments of the Alzheimers peptide beta/A4. *Biochemistry*, **34**, 724–730.
69. Maries, E., Dass, B., Collier, T. J., Kordower, J. H. & Steece-Collier, K. (2003). The role of alpha-synuclein in Parkinson's disease: insights from animal models. *Nature Rev. Neurosci.* **4**, 727–738.
70. Han, H. Y., Weinreb, P. H. & Lansbury, P. T. (1995). The core Alzheimers peptide Nac forms amyloid fibrils which seed and are seeded by beta-amyloid—is Nac a common trigger or target in neurodegenerative disease. *Chem. Biol.* **2**, 163–169.
71. Du, H. N., Tang, L., Luo, X. Y., Li, H. T., Hu, J., Zhou, J. W. & Hu, H. Y. (2003). A peptide motif consisting of glycine, alanine, and valine is required for the fibrillization and cytotoxicity of human alpha-synuclein. *Biochemistry*, **42**, 8870–8878.
72. El-Agnaf, O. M. A. & Irvine, G. B. (2002). Aggregation and neurotoxicity of alpha-synuclein and related peptides. *Biochem. Soc. Trans.* **30**, 559–565.
73. Giasson, B. I., Murray, I. V. J., Trojanowski, J. Q. & Lee, V. M. Y. (2001). A hydrophobic stretch of 12 amino acid residues in the middle of alpha-synuclein is essential for filament assembly. *J. Biol. Chem.* **276**, 2380–2386.
74. Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S. *et al*. (2001). Intrinsically disordered protein. *J. Mol. Graph.* **19**, 26–59.
75. Der-Sarkissian, A., Jao, C. C., Chen, J. & Langen, R. (2003). Structural organization of alpha-synuclein fibrils studied by site-directed spin labeling. *J. Biol. Chem.* **278**, 37530–37535.
76. Wang, W. X. & Hecht, M. H. (2002). Rationally designed mutations convert *de novo* amyloid-like fibrils into monomeric beta-sheet proteins. *Proc. Natl Acad. Sci. USA*, **99**, 2760–2765.
77. Richardson, J. S. & Richardson, D. C. (2002). Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc. Natl Acad. Sci. USA*, **99**, 2754–2759.
78. Otzen, D. E. & Oliveberg, M. (1999). Salt-induced detour through compact regions of the protein folding landscape. *Proc. Natl Acad. Sci. USA*, **96**, 11746–11751.
79. Kruger, R., Kuhn, W., Muller, T., Woitalla, D., Graeber, M., Kosel, S. *et al*. (1998). Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease. *Nature Genet.* **18**, 106–108.
80. Polymeropoulos, M. H., Lavedan, C., Leroy, E., Ide, S. E., Dehejia, A., Dutra, A. *et al*. (1997). Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science*, **276**, 2045–2047.
81. Conway, K. A., Harper, J. D. & Lansbury, P. T. (1998). Accelerated *in vitro* fibril formation by a mutant alpha-synuclein linked to early-onset Parkinson disease. *Nature Med.* **4**, 1318–1320.
82. Saha, A. R., Hill, J., Utton, M. A., Asuni, A. A., Ackerley, S., Grierson, A. J. *et al*. (2004). Parkinson's disease alpha-synuclein mutations exhibit defective axonal transport in cultured neurons. *J. Cell Sci.* **117**, 1017–1024.
83. Zarranz, J. J., Alegre, J., Gomez-Esteban, J. C., Lezcano, E., Ros, R., Ampuero, I. *et al*. (2004). The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia. *Ann. Neurol.* **55**, 164–173.
84. Choi, W., Zibaee, S., Jakes, R., Serpell, L. C., Davletov, B., Crowther, R. A. & Goedert, M. (2004). Mutation E46K increases phospholipid binding and assembly into filaments of human alpha-synuclein. *FEBS Letters*, **576**, 363–368.
85. Ueda, K., Saitoh, T. & Mori, H. (1994). Tissue-dependent alternative splicing of messenger-Rna For Nacp, the precursor of non-A-beta component of Alzheimer's-disease amyloid. *Biochem. Biophys. Res. Commun.* **205**, 1366–1372.
86. Beyer, K., Lao, J. I., Carrato, C., Mate, J. L., Lopez, D., Ferrer, I. & Ariza, A. (2004). Differential expression of alpha-synuclein isoforms in dementia with Lewy bodies. *Neuropathol. Appl. Neurobiol.* **30**, 601–607.
87. Murray, I. V. J., Giasson, B. I., Quinn, S. M., Koppaka, V., Axelsen, P. H., Ischiropoulos, H. *et al*. (2003). Role of alpha-synuclein carboxy-terminus on fibril formation in vitro. *Biochemistry*, **42**, 8530–8540.
88. Serpell, L. C., Berriman, J., Jakes, R., Goedert, M. & Crowther, R. A. (2000). Fiber diffraction of synthetic alpha-synuclein filaments shows amyloid-like cross-beta conformation. *Proc. Natl Acad. Sci. USA*, **97**, 4897–4902.
89. Dedmon, M. M., Lindorff-Larsen, K., Christodoulou, J., Vendruscolo, M. & Dobson, C. M. (2005). Mapping long-range interactions in alpha-synuclein using spin-label NMR and ensemble molecular dynamics simulations. *J. Am. Chem. Soc.* **127**, 476–477.
90. Gamblin, T. C., Berry, R. W. & Binder, L. I. (2003). Modeling Tau polymerization *in vitro*: a review and synthesis. *Biochemistry*, **42**, 15009–15017.
91. Ingram, E. M. & Spillantini, M. G. (2002). Tau gene mutations: dissecting the pathogenesis of FTDP-17. *Trends Mol. Med.* **8**, 555–562.
92. Goedert, M., Spillantini, M. G., Jakes, R., Rutherford, D. & Crowther, R. A. (1989). Multiple isoforms of human microtubule-associated protein-Tau—sequences and localization in neurofibrillary tangles of Alzheimers disease. *Neuron*, **3**, 519–526.
93. Wittmann, C. W., Wszolek, M. F., Shulman, J. M.,

Salvaterra, P. M., Lewis, J., Hutton, M. & Feany, M. B. (2001). Tauopathy in *Drosophila*: neurodegeneration without neurofibrillary tangles. *Science*, **293**, 711–714.
94. Fath, T., Eidenmuller, J. & Brandt, R. (2002). Tau-mediated cytotoxicity in a pseudohyperphosphorylation model of Alzheimer's disease. *J. Neurosci.* **22**, 9733–9741.
95. Goedert, M., Ghetti, B. & Spillantini, M. G. (2000). Tau gene mutations in frontotemporal dementia and parkinsonism linked to chromosome 17 (FTDP-17). Their relevance for understanding the neurogenerative process. *Ann. NY Acad. Sci.* **920**, 74–83.
96. von Bergen, M., Friedhoff, P., Biernat, J., Heberle, J., Mandelkow, E. M. & Mandelkow, E. (2000). Assembly of tau protein into Alzheimer paired helical filaments depends on a local sequence motif ((306)VQIVYK(311)) forming beta structure. *Proc. Natl Acad. Sci. USA*, **97**, 5129–5134.
97. Schenk, D. (2002). Amyloid-beta immunotherapy for Alzheimer's disease: the end of the beginning. *Nature Rev. Neurosci.* **3**, 824–828.
98. Soto, C. (2003). Unfolding the role of protein misfolding in neurodegenerative diseases. *Nature Rev. Neurosci.* **4**, 49–60.
99. Dobson, C. M. (2004). In the footsteps of alchemists. *Science*, **304**, 1259.
100. Dumoulin, M., Last, A. M., Desmyter, A., Decanniere, K., Canet, D., Larsson, G. R. *et al*. (2003). A camelid antibody fragment inhibits the formation of amyloid fibrils by human lysozyme. *Nature*, **424**, 783–788.

Edited by P. T. Lansbury Jr

(Received 10 February 2005; received in revised form 11 April 2005; accepted 11 April 2005)