12. Wang, C. *et al.* TAK1 is a ubiquitin-dependent kinase of MKK and IKK. *Nature* **412,** 346–351 (2001).
13. Shi, C. S. & Kehrl, J. H. TNF-induced GCKR and SAPK activation depends upon the E2/E3 complex Ubc13–Uev1A/TRAF2. *J. Biol. Chem.* **278,** 15429–15434 (2003).
14. Maniatis, T. A ubiquitin ligase complex essential for the NF-κB, Wnt/Wingless, and Hedgehog signaling pathways. *Genes Dev.* **13,** 505–510 (1999).
15. Pomerantz, J. L. & Baltimore, D. Two pathways to NF-κB. *Mol. Cell* **10,** 693–695 (2002).
16. Chu, Z. L., DiDonato, J. A., Hawiger, J. & Ballard, D. W. The tax oncoprotein of human T-cell leukemia virus type 1 associates with and persistently activates IκB kinases containing IKKα and IKKβ. *J. Biol. Chem.* **273,** 15891–15894 (1998).
17. Mercurio, F. *et al.* IKK-1 and IKK-2: cytokine-activated IκB kinases essential for NF-κB activation. *Science* **278,** 860–866 (1997).
18. McManus, M. T. & Sharp, P. A. Gene silencing in mammals by small interfering RNAs. *Nature Rev. Genet.* **3,** 737–747 (2002).
19. Brummelkamp, T. R., Bernards, R. & Agami, R. A system for stable expression of short interfering RNAs in mammalian cells. *Science* **296,** 550–553 (2002).
20. Borodovsky, A. *et al.* Chemistry-based functional proteomics reveals novel members of the deubiquitinating enzyme family. *Chem. Biol.* **9,** 1149–1159 (2002).
21. Karin, M., Cao, Y., Greten, F. R. & Li, Z. W. NF-κB in cancer: from innocent bystander to major culprit. *Nature Rev. Cancer* **2,** 301–310 (2002).
22. Döffinger, R. *et al.* X-linked anhidrotic ectodermal dysplasia with immunodeficiency is caused by impaired NF-κB signaling. *Nature Genet.* **27,** 277–85 (2001).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements We thank R. Agami for the pSUPER vector and advice; R. Beyaert, C. J. Kirschning, M. Krönke, F. Mercurio, K.-I. Nakayama, M. Rowe and M. Treier for plasmids; T. Goncharov, S. Leu, D. Landstein, M. Pasparakis, F. Agou and S. Yamaoka for discussions and support; and I. Beilis and T. Lopez for technical assistance. This work was supported in part by grants from Inter-Lab Ltd, from Ares Trading SA and from the Alfred and Ann Goldstein Foundation to A.K., G.Ca amd D.W., from PTR Pasteur/Necker to G.Co, and from 'La ligue Nationale contre le Cancer' (équipe labellisée) to A.I. A.K. was supported by a postdoctoral fellowship from 'La ligue Nationale contre le Cancer'.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.W. (David.Wallach@weizmann.ac.il) or G.C. (gmcourt@pasteur.fr).

Rationalization of the effects of mutations on peptide and protein aggregation rates

Fabrizio Chiti[1]*, Massimo Stefani[2], Niccolò Taddei[2], Giampietro Ramponi[2] & Christopher M. Dobson[1]

[1]*Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK*
[2]*Dipartimento di Scienze Biochimiche, Università degli Studi di Firenze, Viale Morgagni 50, 50134 Firenze, Italy*

* Present address: Dipartimento di Scienze Biochimiche, Università degli Studi di Firenze, Viale Morgagni 50, 50134 Firenze, Italy

In order for any biological system to function effectively, it is essential to avoid the inherent tendency of proteins to aggregate and form potentially harmful deposits[1–4]. In each of the various pathological conditions associated with protein deposition, such as Alzheimer's and Parkinson's diseases, a specific peptide or protein that is normally soluble is deposited as insoluble aggregates generally referred to as amyloid[2,3]. It is clear that the aggregation process is generally initiated from partially or completely unfolded forms of the peptides and proteins associated with each disease. Here we show that the intrinsic effects of specific mutations on the rates of aggregation of unfolded polypeptide chains can be correlated to a remarkable extent with changes in simple physicochemical properties such as hydrophobicity, secondary structure propensity and charge. This approach allows the pathogenic effects of mutations associated with known familial forms of protein deposition diseases to be rationalized, and more generally enables prediction of the effects of mutations on the aggregation propensity of any polypeptide chain.

The ability to form highly organized aggregates such as the fibrils and plaques associated with the amyloid diseases has been suggested to be a generic property of polypeptide chains, and not simply a feature of the small numbers of proteins so far associated with recognized pathological conditions[5]. Nevertheless, the propensity of a given polypeptide chain to aggregate under specific conditions varies dramatically with its composition and sequence. The proposed generic nature[5] and well established common structural characteristics of aggregates formed from proteins without detectable sequence or structural similarity[6] have encouraged us to investigate whether the effects of amino acid substitutions on the propensity of proteins to aggregate can be described by relatively simple principles that are of general validity. When a protein that is normally folded starts to aggregate, it does so from at least partially unfolded states that are present during or immediately following biosynthesis, or as the result of cellular stress or proteolytic degradation[2,3].

Destabilization of the native state that results in an increased population of partially folded molecules is well established as an important factor in the pathogenic effects of mutations, when the diseases are associated with the deposition of proteins that are globular in their normal functional states[2]. Nevertheless, this factor is not sufficient to explain the pathogenic effects of all mutations associated with such proteins[7,8], in part because the mutations also affect the properties of their aggregation-prone unfolded or partially unfolded states. Clarifying such additional factors that modulate aggregation from non-native states is of fundamental importance for a complete understanding of the relationship between mutation and aggregation behaviour for these systems. More importantly, however, the peptides and proteins

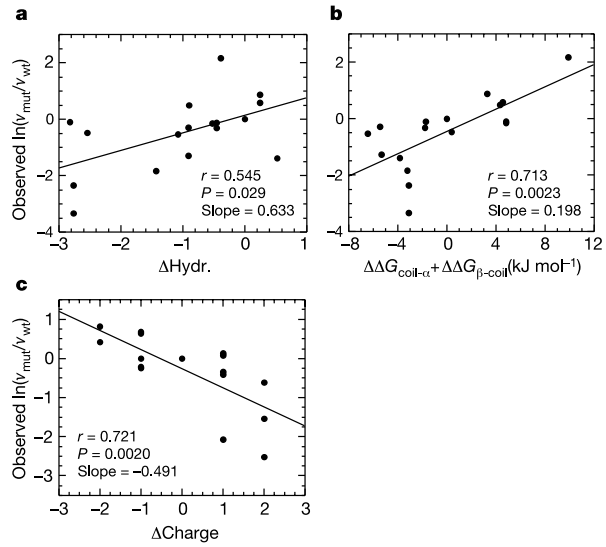

Figure 1 Dependence of the aggregation rate on different physicochemical parameters. Change of the aggregation rate of AcP resulting from mutation plotted against **a**, the predicted change of hydrophobicity, **b**, propensity to convert from an α-helical to a β-sheet conformation, and **c**, charge. All aggregation rate measurements were carried out under conditions in which all the protein variants are substantially unstructured. The mutations reported in **a** and **b**, described previously[9], involve residues 16–31 and 87–98 and do not involve change of charge. The mutations reported in **c**, also described previously[10], were designed to minimize change of hydrophobicity and secondary structure propensities. Values of ΔHydr., ($\Delta\Delta G_{coil\text{-}\alpha} + \Delta\Delta G_{\beta\text{-}coil}$) and ΔCharge were calculated for each mutation as described in Table 1 legend.

© 2003 Nature Publishing Group

Petitioner Merck, Ex. 1157, p. 805
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00052

that aggregate in the most prevalent and important neurodegenerative protein deposition disorders, such as Alzheimer's and Parkinson's diseases, as well as some of the non-neuropathic amyloidoses, have been found to adopt predominantly unstructured conformations in their normal biological state (ref. 3 and references therein). For all these systems, aggregation does not need to be preceded by the unfolding of a globular structure, but involves direct self-assembly of the peptide or protein involved from an ensemble of unstructured conformations.

In the light of the importance of aggregation from at least relatively unstructured states of polypeptide chains, the starting point of our approach has been a detailed mutational study of the aggregation process of a small globular protein under conditions in which the native states of all the studied mutational variants are unstable and at most only marginally populated. Human muscle acylphosphatase (AcP) is highly amenable to such studies, as its aggregation behaviour is well defined and extremely reproducible[9,10]. In particular, the rate of aggregation of AcP from an ensemble of denatured conformations can readily be followed by using a variety of spectroscopic probes, and has now been measured for over 50 mutational variants of this protein[9,10]. By using conditions under which the protein is denatured, changes in aggregation rates can be attributed directly to the intrinsic effects of the amino acid substitutions on the process of self-assembly, without the complications of additional contributions arising from the destabilization of the native state. Many of the mutations examined, particularly those involving residues in two key regions, 16–31 and 87–98, were found to perturb the aggregation rate of AcP very significantly[9,10]. In order to analyse the results, we have characterized all the amino acid substitutions in terms of their effects on three parameters that we anticipated to be important in influencing aggregation rates. These are the change in the hydrophobicity of the polypeptide chain resulting from mutation (ΔHydr.), in the propensity to convert from α-helical to β-sheet structure ($\Delta\Delta G_{\text{coil-}\alpha} + \Delta\Delta G_{\beta\text{-coil}}$) and in

Table 1 **Changes resulting from single-point mutations of unstructured peptides or natively unfolded proteins**

Mutation	ΔHydr. (kcal mol^{-1})*	$\Delta\Delta G_{\beta\text{-coil}}$ (kJ mol^{-1})†	$\Delta\Delta G_{\text{coil-}\alpha}$ (kJ mol^{-1})‡	ΔCharge§	Calculated ln (ν_{mut}/ν_{wt})‖	Observed ln (ν_{mut}/ν_{wt})¶	ref. #
Amylin							
N22A	2.30	−0.95	−3.36	0	0.60	0.70 ± 0.60	13
F23A	−1.88	−4.64	−3.90	0	−2.88	−2.65 ± 0.50	13
G24A	0.39	1.77	−2.84	0	0.04	−0.05 ± 0.30	13
I26A	−1.43	−5.05	−0.32	0	−1.97	−2.40 ± 0.50	13
L27A	−1.43	−2.05	0.36	0	−1.24	−0.95 ± 0.50	13
S20G	1.24	−4.09	0.00	0	−0.03	1.00 ± 0.40	14
Prion peptides							
H111A	3.26	−1.36	−3.21	−1	1.65	0.60 ± 0.70	15
H111K	0.10	0.41	−1.72	0	−0.20	−0.25 ± 0.10	15
A117V☆	0.91	4.63	2.37	0	1.96	1.50 ± 0.70	15
V210I☆	0.52	0.41	−0.97	0	0.22	0.85 ± 0.10	16
α-Synuclein							
A53T☆	−1.39	5.59	2.83	0	0.79	1.20 ± 0.40	17
A76E	−3.30	1.64	0.00	1	−2.25	−2.70 ± 0.70	12
A76R	−4.34	1.64	0.64	−1	−1.80	−0.90 ± 0.40	12
Amyloid β-peptide**							
A21G☆	−0.39	−1.77	3.27	0	0.05	−0.07 ± 0.30	18
E22K☆	0.14	0.14	−1.72	−2	0.76	0.90 ± 0.40	19
E22Q☆	1.61	0.14	0.00	−1	1.54	2.90 ± 0.80	18,19
E22G☆	2.91	−3.41	4.30	−1	2.51	2.05 ± 0.30	20
D23N☆	1.90	4.36	−1.72	−1	2.22	3.95 ± 0.80	18
F19T	−3.27	0.95	−1.76	0	−2.23	−2.50 ± 0.20	21
Tau							
R5L☆	5.77	0.40	−1.00	−1	4.02	3.05 ± 1.50	22
G272V☆	1.30	6.41	−1.71	0	1.75	1.05 ± 0.50	23,24
R406W☆	6.08	1.50	0.00	−1	4.64	1.25 ± 0.80	23–25
Y310W	0.66	−1.77	0.00	0	0.07	0.05 ± 0.05	26
Leucine-rich repeat							
D24N	1.90	4.36	−3.43	−1	1.88	2.10 ± 1.00	27
D24Q	2.51	5.18	−3.10	−1	2.49	1.25 ± 0.70	27
Model peptide							
D6E	0.90	5.04	−2.27	0	1.12	0.40 ± 0.30	28
D6N	1.90	4.36	0.00	1	1.57	0.50 ± 0.30	28

*Calculated using ΔHydr. = Hydr.$_{wt}$ − Hydr.$_{mut}$, where ΔHydr. is the change of hydrophobicity resulting from mutation, and Hydr.$_{wt}$ and Hydr.$_{mut}$ are the hydrophobicity values of the wild-type and mutant residues, respectively (the values of hydrophobicity for all 20 amino acids are listed in the table provided in Supplementary Information 1).

†Calculated using $\Delta\Delta G_{\beta\text{-coil}} = 13.64\,(P_\beta^{wt} - P_\beta^{mut})$ where $\Delta\Delta G_{\beta\text{-coil}}$ is the change of free energy change for the transition random coil → β-sheet resulting from mutation, P_β^{wt} and P_β^{mut} are the β-sheet propensities of the wild-type and mutant residue, respectively (the values of β-sheet propensity for all 20 amino acids are listed in the table provided in Supplementary Information 1), and 13.64 is the conversion constant from the normalized scale to units of kJ mol^{-1}.

‡Calculated using $\Delta\Delta G_{\text{coil-}\alpha} = RT\ln(P_\alpha^{wt}/P_\alpha^{mut})$, where $\Delta\Delta G_{\text{coil-}\alpha}$ is the predicted change of free energy change for the transition α-helix → random coil resulting from mutation, P_α^{wt} and P_α^{mut} are the predicted α-helical propensities (helix percentages) of the wild-type and mutated sequences at the site of mutation, respectively (calculated using the AGADIR algorithm at www.embl-heidelberg.de/Services/serrano/agadir/agadir-start.html); R=0.008314 kJ mol^{-1} K^{-1}.

§Calculated using ΔCharge = |Charge$_{mut}$| − |Charge$_{wt}$|, where ΔCharge is the change of charge resulting from the mutation, and |Charge$_{mut}$| and |Charge$_{wt}$| are the absolute values of charge for the mutated and wild-type sequences, respectively. The operator of 'absolute value' is introduced so that a negative value of ΔCharge results from the equation when the mutation causes the entire protein or peptide to approach neutrality, regardless of the initial sign of charge of the wild-type protein. A positive value of ΔCharge is obtained when the mutation causes the entire protein sequence to deviate further from neutrality.

‖Predicted change of aggregation rate resulting from mutation, calculated using equation (1) (see text for details).

¶Observed change of aggregation rate resulting from mutation (obtained from published experimental data as referenced in the last column of the table). Salt concentrations were within physiological values in the various experiments reported in the table.

#References from which the values of observed ln(ν_{mut}/ν_{wt}) are taken.

☆Natural mutations associated with familial diseases.

**Previously published mutations involving substitutions to proline could not be included in this analysis because of the difficulty in calculating the β-sheet propensities of proline residues from experimental data[29]. Other mutations were also not considered because only qualitative measurements of the effect of such mutations on the aggregation rate were reported. Good agreement is, however, found between our predictions and the qualitative data determined experimentally.

© 2003 Nature Publishing Group

Petitioner Merck, Ex. 1157, p. 806
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00052

the overall charge (ΔCharge). These changes can be quantified using published values of these parameters for all the amino acid residues (reported in Supplementary Information) and the functions given in Table 1 legend.

The change of aggregation rate of AcP upon mutation ($\ln(\nu_{mut}/\nu_{wt})$; see Methods for details) was found to correlate significantly with each of these parameters (Fig. 1a–c). In each of the three plots, however, the data points are considerably scattered around the lines representing the best fits to linear functions. This scatter can readily be attributed to the fact that only a single parameter is considered in each case, to the limitations in the presently available data for predicting accurately changes in the hydrophobicity and secondary structure propensities, and to the varying relative importances of the different sites of mutation in the aggregation process. Despite the scatter present in each plot, however, all three correlations are statistically significant, allowing the mean dependencies of $\ln(\nu_{mut}/\nu_{wt})$ on these three parameters to be calculated. These mean values are represented by the slopes of the lines of best fit resulting from the analysis.

On the assumption that the three factors identified as influencing aggregation are independent of each other and additive, we used a combined function to predict the effect of a mutation on aggregation rate:

$$\ln(\nu_{mut}/\nu_{wt}) = A\Delta\text{Hydr.} + B(\Delta\Delta G_{coil\text{-}\alpha} + \Delta\Delta G_{\beta\text{-}coil}) + C\Delta\text{Charge} \quad (1)$$

Our best present estimates of the A, B and C factors are the slopes of the three plots reported in Fig. 1 (that is, the dependencies of $\ln(\nu_{mut}/\nu_{wt})$ on the three parameters). These are 0.633, 0.198 and −0.491, for A, B and C, respectively.

This function was then used to calculate the changes of the aggregation rate resulting from mutations in a range of other peptides and proteins, particularly those associated with neurodegenerative protein deposition diseases. In order to test this approach, we considered mutations that complied with the following criteria: (1) the experimental value of $\ln(\nu_{mut}/\nu_{wt})$ is directly available or can be determined from data reported in the literature, and (2) the effect of the mutation is reported under conditions in which the polypeptide chain is unstructured. The mutations considered in this analysis, together with the corresponding experimental and calculated values of $\ln(\nu_{mut}/\nu_{wt})$, are listed in Table 1. These mutations include both physiologically relevant ones associated with familial forms of protein deposition diseases (identified in Table 1 with the symbol ☆) and other substitutions that have been investigated to address the importance of particular residues in the aggregation of specific systems. As many peptides or proteins that undergo aggregation in protein deposition diseases are substantially unstructured under physiological conditions, most of the experimental data we could find, and are considered here, were obtained under conditions close to physiological. The various peptide and protein systems reported in Table 1 were studied at different concentrations, but by taking the ratio of ν_{mut} to ν_{wt} the data are normalized in such a manner as to allow for this effect and indeed for the variability in solution conditions.

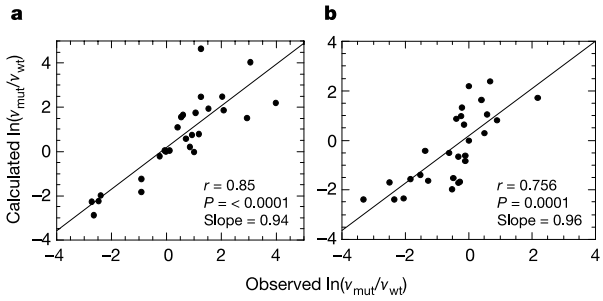

Figure 2 Calculated versus observed changes in aggregation rate upon mutation. **a**, Data relating to mutations of short peptides or natively unfolded proteins, including amylin, amyloid β-peptide, tau and α-synuclein. The 27 mutations shown in the plot and their calculated and experimental $\ln(\nu_{mut}/\nu_{wt})$ values are those listed in Table 1. **b**, Data for the mutations of AcP that are within the two regions of the sequence that have been found to be particularly important for aggregation (encompassing residues 16–31 and 87–98)[9]. All measurements of aggregation rates were carried out under conditions in which all the protein variants are substantially unstructured. In both the analyses (**a** and **b**) the calculated $\ln(\nu_{mut}/\nu_{wt})$ values were obtained using equation (1).

Figure 2a shows the resulting plot of calculated versus experimental values of $\ln(\nu_{mut}/\nu_{wt})$ for the 27 mutations that we identified that satisfied these criteria. The correlation ($r = 0.85$, $P < 0.0001$) is highly significant, and the value of the slope (0.94) is close to 1.0, indicating that there is remarkably good agreement between the calculated and experimental effects of mutations on the aggregation rates. The observed changes of aggregation rate resulting from mutation span a range of nearly 1,000, from 15 times slower to over 50 times faster than the corresponding wild-type polypeptide chains (Fig. 2a and Table 1). Figure 2b shows the calculated versus experimental values of $\ln(\nu_{mut}/\nu_{wt})$ for all the mutations of AcP within the two key regions of the sequence 16–31 and 87–98. The observed correlation is again highly significant ($r = 0.756$ and $P < 0.0001$) and the slope is close to 1.0. It is remarkable that the correlation obtained from mutations of proteins or peptides other than AcP (Fig. 2a) is as significant as that obtained with data from the AcP mutations (Fig. 2b), that is, with data used for the derivation of equation (1).

Examples where close agreement is found between theoretical and experimental values include mutations associated with hereditary spongiform encephalopathies, early-onset Parkinson's disease, early-onset Alzheimer's disease and hereditary cerebral haemorrhage with amyloidosis (Table 1). The effects of mutations that inhibit aggregation, and hence do not cause disease, are also predicted as effectively as those that favour it. Examples include the F19T mutation of the amyloid β-peptide, the A76E mutation of α-synuclein and the L27A mutation of amylin (Table 1). For some mutations, such as the R406W mutation of tau and the V210I mutation of a prion peptide, the agreement is quantitatively less good. However, even in these cases our combined function is generally able to predict qualitatively whether the mutations have an accelerating, decelerating, or only a small effect on the corresponding polypeptides (Table 1).

Interestingly, application of equation (1) reveals that many mutations are predicted not to increase but to reduce very significantly the rates of aggregation of the peptides and proteins that we investigated. It therefore appears that the sequences of natural proteins have not been fully optimized during evolution to resist aggregation, but simply to be sufficiently resistant to aggregation to remain soluble during the normal reproductive life spans[11]. This finding gives rise to the intriguing possibility that there could be a range of mutations within the human population that reduce as well as increase the susceptibility to conditions such as Alzheimer's disease. It will be extremely interesting to test such a prediction, for example by studies of elderly people not afflicted with such illness.

The correlation found in Fig. 2a between the theoretical and experimental effects of mutations on aggregation rates is particularly striking if one considers the heterogeneous groups of peptide and protein systems used in the analysis. Moreover, it is clear that the aggregation rate of a polypeptide chain is more sensitive to mutations occurring within some, often relatively restricted, regions of the sequence[9,12]. The explanation of this latter point is most probably that data reported in the literature are very likely to be those within such regions, as they involve either disease-related

© 2003 Nature Publishing Group

Petitioner Merck, Ex. 1157, p. 807
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00052

mutations or substitutions aimed at identifying such residues in specific experimental studies. Clearly it is important when predicting the effects of mutations either to ensure that they are within such regions or to accept that the predictions could be somewhat higher than the real values.

With these caveats, however, the finding that the effects of mutations on the aggregation rates of these different proteins are closely similar strongly suggests that they result from perturbations to fundamental physicochemical properties of the polymer chains rather than from alterations of specific interactions involving the side chains within the resulting aggregates. This strongly supports the conclusion that there are common principles in the fundamental determinants of aggregation of polypeptide chains despite evident differences in detail[5,30]. Moreover, most of the disease-associated mutations analysed here clearly increase the rate of aggregation (Table 1). This observation indicates that these mutations are likely to be pathogenic primarily as a result of an accelerated rate of aggregation originating from the perturbation of the fundamental parameters that we are now able to identify and describe in a quantitative manner.

The approach described here enables us to begin to identify the factors that affect the intrinsic propensities of polypeptide chains to aggregate. As additional data become available it will undoubtedly be possible to refine further this approach, and also to consider additional parameters that influence the aggregation propensities of peptides and proteins. Even at the present level of sophistication, however, we have shown that a simple function has the ability to rationalize and to predict to a remarkable degree the effects of mutations on the aggregation rates of a wide range of peptides and proteins. As well as giving important insights into the origins of protein misfolding diseases, its use should help substantially in efforts to modify rationally the aggregational properties of proteins or peptides that at present can seriously limit their utilization in research programmes, for medical applications or for biotechnological purposes. □

Methods

The rates of aggregation of the various AcP variants were measured in 25% trifluoroethanol from the time courses of thioflavine T fluorescence, as previously described[9]. The change of aggregation rate as a result of a mutation is expressed in all cases as the natural logarithm of the ratio of the aggregation rate constants of the mutant and wild-type protein ($\ln(\nu_{mut}/\nu_{wt})$).

The experimental values of $\ln(\nu_{mut}/\nu_{wt})$ for mutations of proteins other than AcP were obtained from published data (see last column of Table 1 for references of each mutation). Only data reporting the effect of single-point mutations on the aggregation of proteins and peptides under conditions in which they are unstructured were considered. Data were considered regardless of the experimental techniques used by the different authors to probe aggregation, provided that a quantitative analysis was carried out. When time or rate constants were not explicitly reported, the plots describing the kinetic profiles of aggregation were scanned and computer-analysed to obtain rate constant values. When lag and growth phases were evident in the kinetic profiles of aggregation only the growth phase was considered. When data at fixed periods of time were reported (for example by means of bar graphs), the experimental $\ln(\nu_{mut}/\nu_{wt})$ value was obtained from the ratio of the aggregation parameters obtained from the mutated and wild-type protein/peptide, before equilibrium was reached.

Mutations involving proline residues were not analysed because of the difficulty in obtaining quantitative estimates of the change of β-sheet propensity as a result of these mutations (see Table 1). Nor were mutations considered when substantial discrepancies in the $\ln(\nu_{mut}/\nu_{wt})$ value were reported by different authors (when significant but not substantial discrepancies were present, we considered $\ln(\nu_{mut}/\nu_{wt})$ values resulting from averages of the available data). The reader is referred to Supplementary Information for a detailed description of how the experimental values of $\ln(\nu_{mut}/\nu_{wt})$ were calculated for all mutations.

Received 24 April; accepted 26 June 2003; doi:10.1038/nature01891.

1. Kopito, R. R. Aggresomes, inclusion bodies and protein aggregation. *Trends Cell Biol.* **10,** 524–530 (2000).
2. Kelly, J. W. The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. *Curr. Opin. Struct. Biol.* **8,** 101–106 (1998).
3. Rochet, J. C. & Lansbury, P. T. Jr Amyloid fibrillogenesis: Themes and variations. *Curr. Opin. Struct. Biol.* **10,** 60–68 (2000).
4. Carriò, M. M. & Villaverde, A. Construction and deconstruction of bacterial inclusion bodies. *J. Biotech.* **96,** 3–12 (2002).
5. Dobson, C. M. Protein misfolding, evolution and disease. *Trends Biochem. Sci.* **24,** 329–332 (1999).
6. Sunde, M. & Blake, C. The structure of amyloid fibrils by electron microscopy and X-ray diffraction. *Adv. Protein Chem.* **50,** 123–159 (1997).
7. Liemann, S. & Glockshuber, R. Influence of amino acid substitutions related to inherited human prion diseases on the thermodynamic stability of the cellular prion protein. *Biochemistry* **38,** 3258–3267 (1999).
8. Hammarstrom, P., Wiseman, R. L., Powers, E. T. & Kelly, J. W. Prevention of transthyretin amyloid disease by changing protein misfolding energetics. *Science* **299,** 713–716 (2003).
9. Chiti, F. *et al.* Kinetic partitioning of protein folding and aggregation. *Nature Struct. Biol.* **9,** 137–143 (2002).
10. Chiti, F. *et al.* 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 (2002).
11. Dobson, C. M. Getting out of shape. *Nature* **418,** 729–730 (2002).
12. Giasson, B. I., Murray, I. V., Trojanowski, J. Q. & Lee, V. M. A hydrophobic stretch of 12 amino acid residues in the middle of α-synuclein is essential for filament assembly. *J. Biol. Chem.* **276,** 2380–2386 (2001).
13. Azriel, R. & Gazit, E. 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 (2001).
14. Sakagashira, S. *et al.* S20G mutant amylin exhibits increased in vitro amyloidogenicity and increased intracellular cytotoxicity compared to wild-type amylin. *Am. J. Pathol.* **157,** 2101–2109 (2000).
15. Salmona, M. *et al.* Molecular determinants of the physicochemical properties of a critical prion protein region comprising residues 106–126. *Biochem. J.* **342,** 207–214 (1999).
16. Thompson, A. J., Barnham, K. J., Norton, R. S. & Barrow, C. J. The Val-210-Ile pathogenic Creutzfeldt-Jakob disease mutation increases both the helical and aggregation propensities of a sequence corresponding to helix-3 of PrP(C). *Biochim. Biophys. Acta* **1544,** 242–254 (2001).
17. Conway, K. A. *et al.* Acceleration of oligomerization, not fibrillization, is a shared property of both alpha-synuclein mutations linked to early-onset Parkinson's disease: Implications for pathogenesis and therapy. *Proc. Natl Acad. Sci. USA* **97,** 571–576 (2000).
18. Van Nostrand, W. E., Melchor, J. P., Cho, H. S., Greenberg, S. M. & Rebeck, G. W. Pathogenic effects of D23N Iowa mutant amyloid β-protein. *J. Biol. Chem.* **276,** 32860–32866 (2001).
19. Miravalle, L. *et al.* Substitutions at codon 22 of Alzheimer's abeta peptide induce diverse conformational changes and apoptotic effects in human cerebral endothelial cells. *J. Biol. Chem.* **275,** 27110–27116 (2000).
20. Nilsberth, C. *et al.* The 'Arctic' APP mutation (E693G) causes Alzheimer's disease by enhanced Aβ protofibril formation. *Nature Neurosci.* **4,** 887–893 (2001).
21. Esler, W. P. *et al.* Point substitution in the central hydrophobic cluster of a human β-amyloid congener disrupts peptide folding and abolishes plaque competence. *Biochemistry* **35,** 13914–13921 (1996).
22. Gamblin, T. C., Berry, R. W. & Binder, L. I. Tau polymerization: Role of the amino terminus. *Biochemistry* **42,** 2252–2257 (2003).
23. Barghorn, S. *et al.* Structure, microtubule interactions, and paired helical filament aggregation by tau mutants of frontotemporal dementias. *Biochemistry* **39,** 11714–11721 (2000).
24. Gamblin, T. C. *et al.* In vitro polymerization of tau protein monitored by laser light scattering: Method and application to the study of FTDP-17 mutants. *Biochemistry* **39,** 6136–6144 (2000).
25. Nacharaju, P. *et al.* Accelerated filament formation from tau protein with specific FTDP-17 missense mutations. *FEBS Lett.* **447,** 195–199 (1999).
26. Li, L., Von Bergen, M., Mandelkow, E. M. & Mandelkow, E. Structure, stability, and aggregation of paired helical filaments from tau protein and FTDP-17 mutants probed by tryptophan scanning mutagenesis. *J. Biol. Chem.* **277,** 41390–41400 (2002).
27. Symmons, M. F., Buchanan, S. G., Clarke, D. T., Jones, G. & Gay, N. J. X-ray diffraction and far-UV CD studies of filaments formed by a leucine-rich repeat peptide: Structural similarity to the amyloid fibrils of prions and Alzheimer's disease β-protein. *FEBS Lett.* **412,** 397–403 (1997).
28. Orpiszewski, J. & Benson, M. D. Induction of beta-sheet structure in amyloidogenic peptides by neutralization of aspartate: A model for amyloid nucleation. *J. Mol. Biol.* **289,** 413–428 (1999).
29. Street, A. G. & Mayo, S. L. Intrinsic β-sheet propensities result from van der Waals interactions between side chains and the local backbone. *Proc. Natl Acad. Sci. USA* **96,** 9074–9076 (1999).
30. Bucciantini, M. *et al.* Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases, *Nature* **416,** 507–511 (2002).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements We thank M. Calamai, J. Zurdo and M. Vendruscolo for critical reading of the manuscript. The Dipartimento di Scienze Biochimiche in Florence is supported by the Italian MIUR (Progetto "Genetica Molecolare" and Progetto FIRB "Folding di proteine: l'altra metà del codice genetico"). We are grateful for support from the Wellcome Trust (F.C. and C.M.D.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.M.D. (cmd44@cam.ac.uk).

© 2003 Nature Publishing Group

Petitioner Merck, Ex. 1157, p. 808
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00052

Supplementary information - 1

Scales of hydrophobicity, β-sheet propensity and charge for the 20 natural amino acids			
amino acid residue	hydrophobicity (kcal mol^{-1}) [a]	β-sheet propensity [b]	charge [c]
Arg (R)	3.95	0.35	+1
Lys (K)	2.77	0.34	+1
Asp (D)	3.81	0.72	-1
Glu (E)	2.91	0.35	-1
Asn (N)	1.91	0.40	0
Gln (Q)	1.30	0.34	0
His (H)	0.64 (2.87) [d]	0.37	0 (+1) [d]
Ser (S)	1.24	0.30	0
Thr (T)	1.00	0.06	0
Tyr (Y)	-1.47	0.11	0
Gly (G)	0.00	0.60	0
Pro (P)	-0.99	n.d.	0
Cys (C)	-0.25	0.25	0
Ala (A)	-0.39	0.47	0
Trp (W)	-2.13	0.24	0
Met (M)	-0.96	0.26	0
Phe (F)	-2.27	0.13	0
Val (V)	-1.30	0.13	0
Ile (I)	-1.82	0.10	0
Leu (L)	-1.82	0.32	0

[a] hydrophobicity values of the 20 amino acid residues at neutral pH based on the partition coefficients from water to octanol. The data are from column 6 of Table 4.8 in ref. 30.

[b] β-sheet propensities of the 20 amino acid residues normalised from 0 (high β-sheet propensity) to 1 (low β-sheet propensity). The data are from column 4 of Table 1 of ref. 29. The β-sheet propensity of proline is not reported due to the difficulty in determining it experimentally. The β-sheet propensity of glycine is from theoretical calculations.

[c] values of charge are at neutral pH.

[d] values in brackets are at a pH lower than 6.0, when the histidine residue is positively charged

Petitioner Merck, Ex. 1157, p. 809
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00052

Supplementary information - 2

In this section we describe the way we utilised the experimental data from the literature to determine the experimental values of ln (v_{mut}/v_{wt}) for each of the mutations reported in Table 1. The reader should also refer to the *Methods* section.

---

mutations of amylin

Experimental data of ln (v_{mut}/v_{wt}) were all calculated from Fig. 2B (data points at 4 min) of ref. 13

S20G mutation of amylin

Experimental data of ln (v_{mut}/v_{wt}) on the S20G mutation of amylin were from Fig. 5 of ref. 14. Data were replotted to obtain rate constants within the elongation phases. The value of ln (v_{mut}/v_{wt}) considered in our analysis is the average of the two ln (v_{mut}/v_{wt}) values obtained from the two reported concentrations.

H111A, H111K and A117V mutations of a prion peptide

Experimental data of ln (v_{mut}/v_{wt}) on the 106-126 peptide of the human prion were from Fig. 2 of ref. 15. Data were replotted to determine the initial rates of monomer depletion.

V210I mutation of a prion peptide

Experimental data of ln (v_{mut}/v_{wt}) on the 198-218 peptide of the human prion were from Fig. 8 (aggregation rates were taken from the slopes of the reported plot) of ref. 16

A53T mutation of α-synuclein

Experimental data of ln (v_{mut}/v_{wt}) on the A53T mutation of α-synuclein were from ref. 17: data were taken from figure 1B (time of 14 days), figure 2A (time of 49 days) and from figure 3A (time 66 days). The reported value of ln (v_{mut}/v_{wt}) results from an average of the three values

---

A76E and A76R mutations of α-synuclein

Experimental data of ln (v_{mut}/v_{wt}) on the A76E and A76R mutations of α-synuclein were from Fig. 3 (time of 2 days) of ref. 12

A21G and D23N mutations of Aβ

Experimental data of ln (v_{mut}/v_{wt}) on the A21G and D23N mutations of Aβ were from Fig. 3 of ref. 18: data were replotted to obtain the rate of depletion of Congo red. This implied considering the rate of the first 6 hr for the D23N mutant (before the equilibrium was reached) and between 0 and 48 hr for the A21G and wild-type peptides.

E22K mutation of Aβ

Experimental data of ln (v_{mut}/v_{wt}) on the E22K mutation of Aβ were from Fig. 2 of ref. 19: the data points were replotted and fitted to a single exponential function to obtain rate constant values.

E22Q mutation of Aβ

Experimental data of ln (v_{mut}/v_{wt}) on the E22Q mutation of Aβ were from (1) Fig. 2 of ref. 19: the data points were replotted and fitted to a single exponential function to obtain rate constant values; (2) Fig. 3 of ref. 18: data were replotted to obtain the rate of depletion of Congo red; this implied considering the rate of the first 6 hr for the E22Q mutant (before the equilibrium was reached) and between 0 and 48 hr for the wild-type peptide. The reported value of ln (v_{mut}/v_{wt}) results from an average of the two values.

Petitioner Merck, Ex. 1157, p. 810
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00052

E22G mutation of Aβ

Experimental data of ln (v_{mut}/v_{wt}) on the E22G mutation of Aβ were from Fig. 5a,b of ref. 20: data were replotted to obtain the rate of depletion of monome/dimer. This implied considering the rate of the first 5 hr for the E22G mutant (before the equilibrium was reached) and between 0 and 50 hr for the wild-type peptide.

F19T mutation of Aβ

Experimental data of ln (v_{mut}/v_{wt}) on the F19T mutation of Aβ were from Fig. 4 of ref. 21: similar values of ln (v_{mut}/v_{wt}) are obtained at different Aβ concentrations.

R5L mutation of tau

Experimental data of ln (v_{mut}/v_{wt}) on the R5L mutation of tau were from Fig. 5 of ref. 22: the data points were replotted and fitted to a single exponential function (for the wild-type protein) or a double exponential function (for the mutant) to obtain rate constant values; experimental data of ln(v_{mut}/v_{wt}) were calculated for both phases observed for the mutant. The reported value of ln (v_{mut}/v_{wt}) results from an average of the two values.

G272V mutation of tau

Experimental data of ln (v_{mut}/v_{wt}) on the G272V mutation of tau were from (1) Table 1 (time constants) of ref. 24; (2) Figures 5 and 6 (rates during elongation phases) of ref. 23. The reported value of ln (v_{mut}/v_{wt}) results from an average of the three values

---

R406W mutation of tau

Experimental data of ln (v_{mut}/v_{wt}) on the R406W mutation of tau were from Table 1 (time constants) of ref. 24, Figure 5 (rates during elongation phases) of ref. 23 and from Fig. 1 (rates during elongation phases) of ref. 25. The reported value of ln (v_{mut}/v_{wt}) results from an average of the three values.

Y310W mutation of tau

Experimental data of ln (v_{mut}/v_{wt}) on the Y310W mutation of tau were from Figure 3A of ref. 26: the data points in the presence of heparin were replotted and fitted to single exponential functions to obtain rate constant values.

Mutations of the leucine rich repeat peptide

Experimental data of ln (v_{mut}/v_{wt}) on the leucine-rich repeat were from Fig. 2 ref. 27. The aggregation rates were taken from the slopes of the reported plot

---

mutations of the VTVKVDAVKVTV peptide

Experimental data of ln (v_{mut}/v_{wt}) on the 12 residue model peptide were from Fig. 8 of ref. 28 (mean residue ellipticity at the peak in the 215-220 nm region subtracted by mean residue ellipticity for random coil obtained from fig. 6a).

Petitioner Merck, Ex. 1157, p. 811
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00052