

JMBAvailable online at www.sciencedirect.com ScienceDirect

Competition between Intramolecular and Intermolecular Interactions in an Amyloid-Forming Protein

Katy E. Routledge¹, Gian Gaetano Tartaglia², Geoffrey W. Platt¹,
Michele Vendruscolo^{2*} and Sheena E. Radford^{1*}

¹*Astbury Centre for Structural Molecular Biology, Institute of Molecular and Cellular Biology, University of Leeds, Leeds LS2 9JT, UK*

²*Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK*

Received 18 February 2009;
received in revised form
9 April 2009;
accepted 20 April 2009
Available online
23 April 2009

Despite much progress in understanding the folding and the aggregation processes of proteins, the rules defining their interplay have yet to be fully defined. This problem is of particular importance since many diseases are initiated by protein unfolding and hence the propensity to aggregate competes with intramolecular collapse and other folding events. Here, we describe the roles of intramolecular and intermolecular interactions in defining the length of the lag time and the apparent rate of elongation of the 100-residue protein human β_2 -microglobulin at pH 2.5, commencing from an acid-denatured state that lacks persistent structure but contains significant non-random hydrophobic interactions. Using a combination of site-directed mutagenesis, quantitative kinetic analysis and computational methods, we show that only a single region of about 10 residues in length, determines the rate of fibril formation, despite the fact that other regions exhibit a significant intrinsic propensity for aggregation. We rationalise these results by analysing the effect of incorporating the conformational properties of acid-unfolded β_2 -microglobulin and its variants at pH 2.5 as measured by NMR spectroscopy into the Zygggregator aggregation prediction algorithm. These results demonstrate that residual structure in the precursor state modulates the intrinsic propensity of the polypeptide chain to aggregate and that the algorithm developed here allows the key regions for aggregation to be more clearly identified and the rates of their self-association to be predicted. Given the common propensity of unfolded chains to form non-random intramolecular interactions as monomers and to self-assemble subsequently into amyloid fibrils, the approach developed should find widespread utility for the prediction of regions important in amyloid formation and their rates of self-assembly.

© 2009 Elsevier Ltd. Open access under [CC BY license](http://creativecommons.org/licenses/by/3.0/).

Edited by J. Weissman

Keywords: amyloid; aggregation; NMR relaxation; prediction; unfolded state

Introduction

A wide range of highly debilitating diseases are associated with the failure of proteins to maintain their native structures and remain soluble—an abnormal behaviour that results in their aggregation into fibrillar assemblies.¹ These amyloid fibrils exhibit a common cross- β structure^{2,3} despite the lack of sequence and structural similarities between their precursor proteins.¹ It has been suggested that the

ability to form the characteristic cross- β structure of amyloid is an inherent latent property of polypeptide chains, as many proteins not known to be involved in disease can produce amyloid-like fibrils *in vitro* under appropriate conditions.³ Therefore, studying the determinants of fibril formation using model systems can provide important insights into how proteins assemble into the amyloid fold, with applications for understanding both advantageous^{4,5} and deleterious^{1,6} consequences of this protein self-assembly process.

Whilst insights into the atomic details of cross- β spines of peptides assembled into amyloid-like structures have been obtained,^{7–10} the molecular mechanisms of self-assembly into amyloid fibrils remain poorly understood. One aspect that has been

*Corresponding authors. E-mail addresses:
mv245@cam.ac.uk; S.E.Radford@leeds.ac.uk.

Abbreviations used: β_2 m, β_2 -microglobulin; ThT, thioflavin T.

clarified is that the same fundamental forces that lead to folding, which include hydrogen bonding as well as hydrophobic and electrostatic interactions, are also responsible for aggregation. Indeed, native states play a protective role by sequestering in their interior the most aggregation-prone regions of amino acid sequences.¹¹ As a consequence, fibril formation may be inhibited by increasing the stability of the native state through mutational analysis or the addition of small molecules.^{12,13} Conversely, therefore, the native state must be destabilised to allow intermolecular interactions that are favourable for aggregation to occur.^{11,14–18} Hence, even in the absence of folding to a native conformation, intramolecular interactions including fluctuating side chain–side chain interactions commonly associated with non-native collapsed states could play a role in modulating the aggregation propensity of a polypeptide chain (Fig. 1).

To investigate this hypothesis, we carried out a systematic analysis of the aggregation kinetics at low pH of β_2 -microglobulin (β_2m) and 30 mutational variants. Under these conditions, β_2m forms fibrils with >90% yield and reproducible kinetics,^{19,20} commencing from a precursor that is unfolded but contains significant non-random structure involving two hydrophobic clusters stabilised by the native disulfide bond.²¹ The results show that only a single region involving residues ~60–70 in the full-length, 100-residue protein plays a major role in determining the kinetics of fibril formation, by contrast with some, but not all, predictions inferred from peptide studies^{22–24} and numerical algorithms.^{25–27} By taking into account the presence of non-random residual structure in the sequence of β_2m at low pH provided by R_2 measurements²¹ and its dependence on the protein sequence, we have developed an enhanced prediction method based on the physicochemical properties of the amino acid sequence¹¹ modulated by its propensity to form non-random intramolecular contacts (Fig. 1). We show that this approach extends the model that we proposed previously for folded states¹¹ in providing an accurate description of the competition between intramolecular interactions and aggregation under non-native conditions. Application to the case of β_2m at low pH demonstrates that

this method accurately predicts the regions essential for amyloid formation and in controlling its rate. Given the generic propensity of unfolded chains to collapse and the prevalence of these structures in amyloid precursors of medical importance,²⁸ the method developed should find widespread utility for the prediction of β -aggregation in other systems.

Results

Design of β_2m variants

Under native conditions, β_2m forms a disulfide-bonded immunoglobulin fold (Fig. 2a).³² By contrast, at pH 2.5, β_2m is denatured, forming a non-random collapsed state that lacks ordered secondary structure but contains two non-native hydrophobic clusters (residues 29–51 and 58–79) that are mutually interacting and stabilised by the native disulfide bond.²¹ Under these conditions, β_2m rapidly and spontaneously assembles into fibrils with the morphological, tinctorial, structural and kinetic characteristics of amyloid.^{20,33} Data collected under a variety of different experimental conditions suggest that β_2m is highly amyloidogenic, with more than 60% of the sequence observed to form amyloid-like fibrils when created as peptide fragments^{7,22–24,29,30} (Fig. 2b). The use of single point substitutions and chimeric proteins also suggested that numerous individual residues and/or regions of the protein are involved in fibril formation (see Ref. 34 for a collated list of the mutations of β_2m made to date). These results highlight the potential importance of much of the sequence of β_2m for fibril formation. How this is modulated in the context of the intact protein, however, remains unknown. To address this question, a systematic mutagenesis study is required to resolve residue-specific details of the amyloidogenic determinants of β_2m in the context of the full-length, oxidised protein under a single, identical set of experimental conditions.

In an earlier study, we investigated the influence of just one region of β_2m (residues 60–70, previously

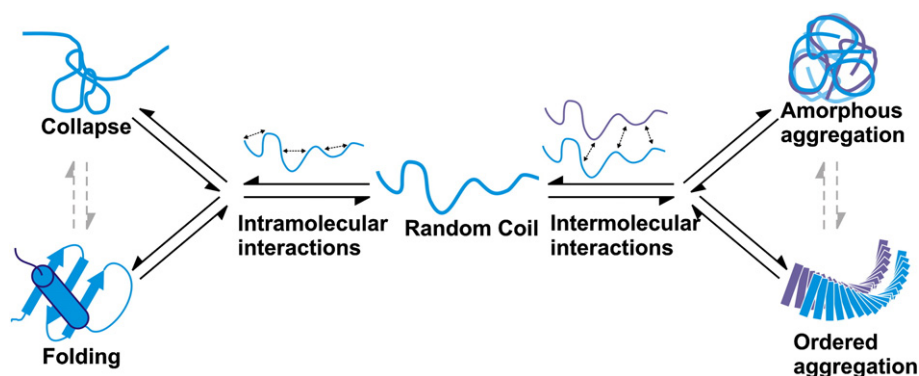


Fig. 1. Schematic illustration of the competition between intramolecular and intermolecular interactions in protein folding and aggregation. When intramolecular interactions prevail, proteins form partially or fully folded states (left pathway). By contrast, when intermolecular interactions dominate, protein aggregation results (right pathway).

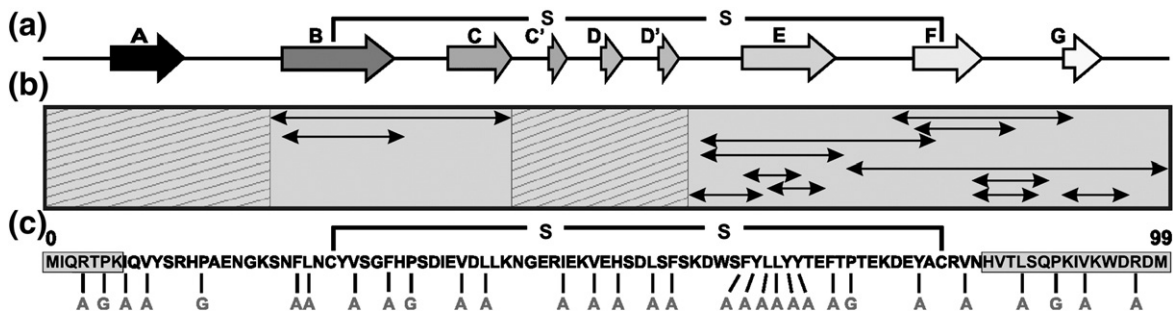


Fig. 2. Sequence and fibrillogenic properties of β_2m and its peptide fragments. (a) Secondary-structure elements of native human β_2m . The native β -strands (arrows) and the position of the disulfide bond are shown. (b) Regions (double-headed arrows) of β_2m that have been shown to form fibrils *in vitro* when created as peptide fragments.^{7,22–24,29–31} Note that different conditions were utilised in different studies. Hashed regions (residues 1–19 and 41–59) have not been shown to form fibrils under any conditions tested thus far. (c) Primary sequence of recombinant wild-type β_2m showing the position and type of amino acid substitutions introduced in this study. The residues at the N- and C-termini truncated to form $\Delta N6$ and $\Delta C83$ are highlighted in grey. The disulfide bond was maintained in all analyses.

suggested as a highly amyloidogenic sequence^{22,26,35}) in determining the rate of aggregation of the full-length protein.¹⁹ Building on these results, we describe here a systematic analysis of the aggregation properties of wild-type β_2m and 30 variants that span its 100-residue sequence, including residues involved in all of the native β -strands or their interconnecting loops (Fig. 2c and Supplementary Table 1). In the majority of the variants created, hydrophobic residues were substituted with alanine. Additionally, the five prolines were individually substituted with glycine. Finally, two truncation variants were created: $\Delta N6$, a species found in *ex vivo* deposits and known to have increased amyloid potential *in vitro*,³⁶ and a novel variant, $\Delta C83$, in which residues 84–99 were removed. The C-terminal 28 residues of β_2m were shown to be highly amyloidogenic when created as a peptide fragment³⁰ and to contain a region (residues 83–89) that determines the vastly different amyloid potentials of murine and human β_2m .³¹

Role of individual residues in the rates of fibril nucleation and elongation

To assess the effect of each mutation on the kinetics of fibril formation, we monitored the length of the lag phase and the apparent rate of elongation using thioflavin T (ThT) fluorescence as a probe of fibrillogenesis, either in the presence or in the absence of fibril seeds (Fig. 3). Conditions that enable reliable measurement of the rate of fibril formation over multiple samples recorded simultaneously were used (Materials and Methods).¹⁹ Using this procedure, we determined the mean and variance of the lag time and apparent elongation rate of each variant (Fig. 3). The kinetics of fibril formation of wild-type β_2m were recorded concurrently with each variant and used to normalise the data obtained. With the exception of the variant $\Delta C83$, all proteins formed fibrils in unseeded reactions within 72 h, each resulting in fibrils that are morphologically indistinguishable by negative-stain electron microscopy from those formed from the wild-type protein (Fig. 4a). Under different incubation

conditions (Materials and Methods), $\Delta C83$ also formed long, straight fibrils (Fig. 4a). The lag time of $\Delta C83$ aggregation, therefore, was not compared directly with the other variants. The data nonetheless suggest that the presence of the C-terminal region is important, but not critical, in the initial stages of fibril formation. For the other mutant proteins, the effect of sequence variation on the lag time is dependent on the site of substitution and the type of modification introduced. In particular, five of the variants tested (H51A, F62A, Y63A, L65A and V82A) showed marked differences (>2- to 7-fold changes) in the average lag time compared with that of wild-type β_2m , whilst for the other 24 variants, the lag time was relatively unperturbed (resulting in ≤ 2 -fold change in the lag time upon mutation) (Fig. 3b and Supplementary Table 1). All the P to G variants showed little difference in lag time compared with wild-type β_2m (<2-fold change) despite the fact that substitution of Pro with Gly probes the effects of altering both the main chain and side chain in determining aggregation (Fig. 3b and Supplementary Table 1). Interestingly, L65A has the greatest effect on the lag time, whilst identical amino acid substitutions elsewhere in the sequence result in small (<2-fold) changes (e.g., compare the data for L65A with L87A) (Fig. 3b and Supplementary Table 1). The only variant to have a significantly shortened lag time compared with wild-type β_2m was V82A. Why no other residue showed such behaviour is not clear: further substitutions would be needed to explore this further. Together, the data show that for this 100-residue protein, only residues 51, 62, 63 and 65 appear to be critically involved in determining the lag time of fibril formation, with other residues, including those in the C-terminal half of the polypeptide chain, playing a supporting role. This observation contrasts with the behaviour of peptide fragments, which suggested that three regions (residues 20–40, 60–70 and ~80–100) are important for fibril formation^{7,22,29,30} (Fig. 2b). Furthermore, the region of the protein containing His51 has not been observed to be prone to fibril formation as

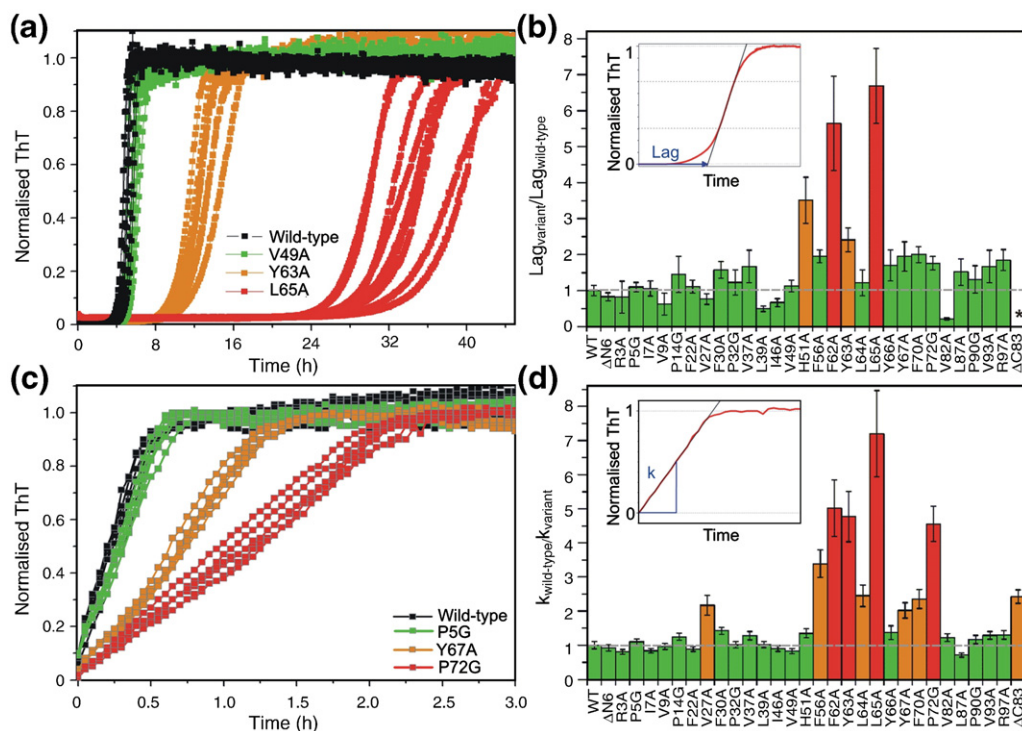


Fig. 3. Comparison of the fibrillation kinetics of different β_2m variants compared with the wild-type protein. Non-seeded (a) and seeded (c) fibril growth kinetics of wild-type β_2m and representative variants measured using ThT fluorescence. Assays were carried out as previously described (Materials and Methods).¹⁹ (b) Normalised lag times ($Lag_{variant}/Lag_{wild\ type}$) determined in unseeded growth experiments. The variants are grouped according to changes in lag time compared with wild-type β_2m ; <2-fold change, green; 2- to 4-fold change, orange; >4-fold change, red. *Under these conditions, fibrils were not formed within the time frame of these experiments (72 h). Inset: illustration of the method used to calculate the lag time. (d) Relative apparent elongation rate ($k_{wild\ type}/k_{variant}$) determined from seeded growth experiments (see Materials and Methods). The data are coloured as in (b). Inset: illustration of the method used to calculate the apparent elongation rate (k). Error bars are ± 1 SD.

part of an isolated fragment, yet this residue appears to play a significant role in determining the length of the lag phase in the conditions tested. These results indicate that there are fundamental differences in aggregation potential of particular residues or certain sequences when isolated as peptides compared with their behaviour when included within the context of the full-length protein.

Seeds formed by fragmenting fibrils prepared from the wild-type protein at pH 2.5 were next added to each reaction to determine the apparent elongation rate of each variant from a common template (Materials and Methods). Under these conditions, highly reproducible kinetics ensue, allowing accurate determination of the apparent elongation rates¹⁹ and a more detailed comparison of the influence of different amino acid substitutions on the aggregation kinetics. All variants were able to extend seeds made from wild-type β_2m , consistent with each variant assembling into a common fibrillar architecture (Fig. 4b). These experiments revealed that variants located predominantly in one region of the sequence (residues ~60–70) reduce the apparent elongation rate with respect to the wild-type protein (Fig. 3c and d and Supplementary Table 1), highlighting the key role of this region in determining both the length of the lag

time and the apparent rate of elongation. Three further variants, V27A, F56A and P72G, also affected the apparent elongation rate, suggesting that these residues are also important in the elongation process.

The combined results from both sets of experimental data demonstrate the remarkable and unexpected result that only one region of contiguous amino acids (residues ~60–70) within intact denatured β_2m plays a pivotal role in determining the rate of fibril nucleation and elongation at low pH. This finding is in marked contrast with predictions of intrinsically amyloidogenic regions based on the physicochemical properties of the amino acid sequence^{25–27} and experimental analyses using peptide fragments (Fig. 2b). Akin to a protein folding reaction,³⁷ therefore, the kinetics of fibril formation seem to be regulated by a few key residues, with the rest of the sequence providing a scaffold for encouraging profitable interactions.

Non-random structure within denatured β_2m modulates its aggregation propensity

The ability to form fibrils is thought to be a common property of most polypeptide chains, as long as suitable environmental conditions are found.³ Conditions that

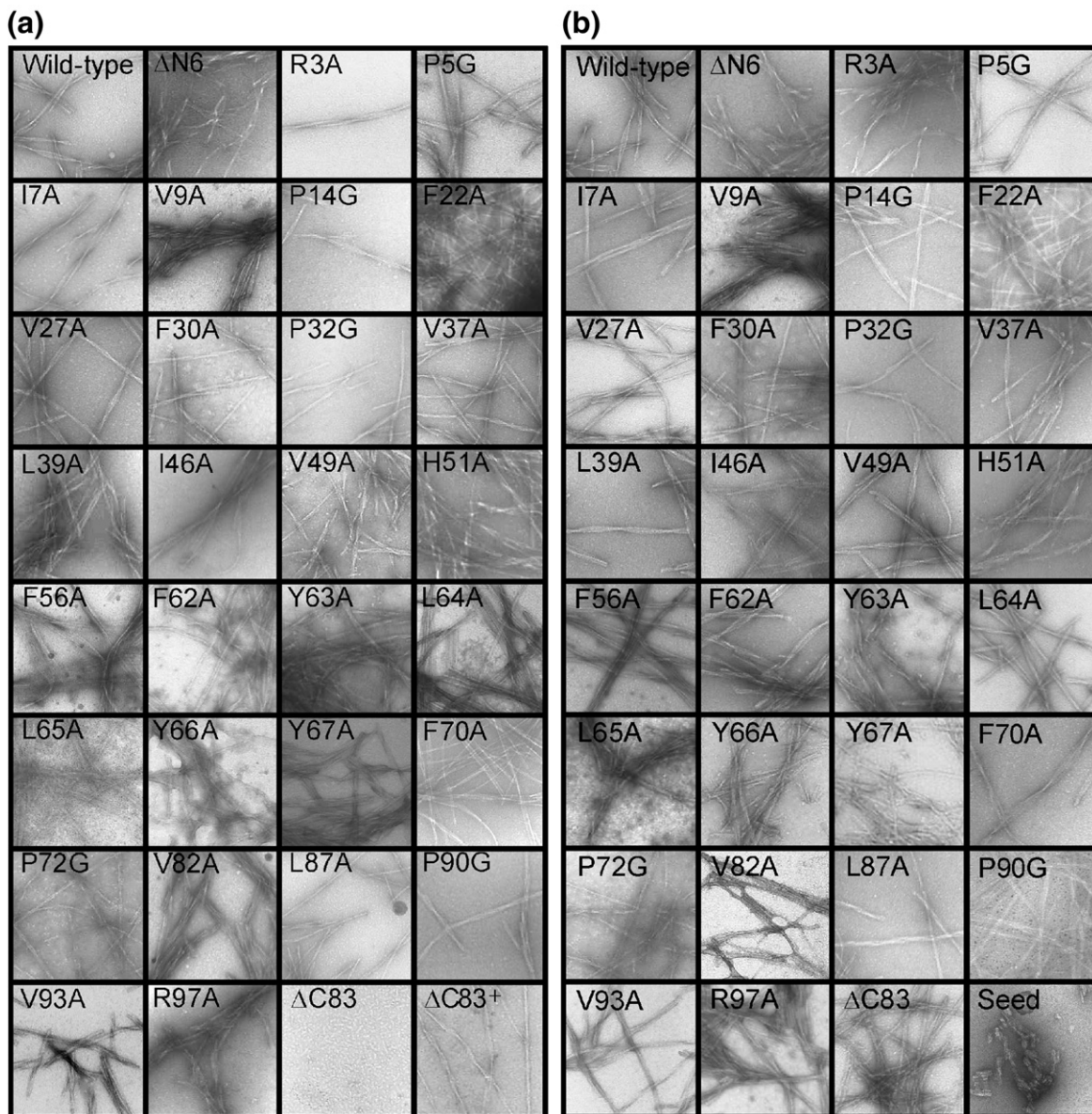


Fig. 4. Negatively stained electron micrographs of the end products of fibril formation for wild-type β_2m and 30 variants. (a) End products of non-seeded fibril formation (72 h). $\Delta C83$ fibrils were generated under different conditions from the other variants ($\Delta C83^+$) (Materials and Methods). (b) End products of seeded growth of each variant from wild-type seeds. Each square measures $0.5 \mu m \times 0.5 \mu m$.

result in the exposure of the polypeptide chain to solvent thus reveal its inherent aggregation propensity, a property that is dependent on the intrinsic physicochemical properties of the amino acids in the sequence.³ These properties have been used as a basis for the formulation of phenomenological models capable of predicting changes in aggregation propensity upon mutation, as well as the absolute rates of aggregation of peptides and unstructured polypeptide chains.^{26,27,35,38} Predictions of the aggregation-prone regions of β_2m using TANGO, Aggrescan and Zyggregator^{25–27} identified the regions ~20–40 and ~50–70 as having high intrinsic aggregation propensities (Figs. 5 and 6a), albeit with differences in the magnitude of the aggregation propensity for each region. The prediction of residues ~50–70 as the most aggregation-prone region is in agreement with

the experimental data presented here (Fig. 3b and d and Supplementary Table 1) and with those previously presented.^{19,22} Interestingly, of the six residues mutated in region 20–40 (F22A, V27A, F30A, P32G, V37A and L39A), a sequence that is predicted to have a high intrinsic aggregation propensity by two of the algorithms (Aggrescan and Zyggregator) (Figs. 5 and 6a) and supported using peptide studies^{23,29} (Fig. 2b), only one mutation (V27A) caused a change (~2-fold) in the apparent elongation rate, and none significantly affected the lag time (Fig. 3b and d). It is therefore important to understand why the intrinsic aggregation propensity is not by itself a good predictor of the tendency to aggregate for this region of β_2m . In the case of fully folded proteins, we already demonstrated that regions that have a high intrinsic aggregation propensity might

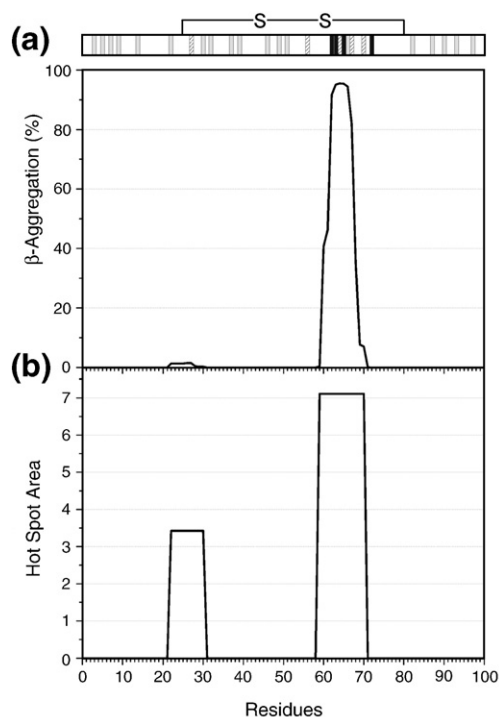


Fig. 5. The aggregation propensity of the primary sequence of β_2m predicted by the algorithms (a) TANGO²⁶ and (b) Aggrescan.²⁷ A schematic illustration of the effects of amino acid substitution on the rates of fibril elongation determined here is shown above. The effects are classified into three groups dependent on the change in apparent elongation rate; >2-fold change, grey; intermediate, 2- to 4-fold change, hashed; slow, >4-fold change, black.

not play an important role in the aggregation process because they have also a high propensity to form intramolecular interactions in the native structure and, hence, be protected, at least in part, from giving rise to aggregation.¹¹ We therefore investigated here whether intramolecular interactions are also able to alter the intrinsic aggregation propensities under non-native conditions. In the case of β_2m , the region comprising residues ~20–40, which has an intrinsic aggregation propensity predicted by Zygggregator similar to that of residues ~50–70 (Fig. 6a), not only contains significant non-native structure in the acid-denatured state of β_2m ^{19,21} but also contains Cys25, a residue that forms a disulfide bond with Cys80 that is essential both for the formation of a stable native structure and the generation of amyloid fibrils from full-length β_2m .^{39,40} Given that fragment 20–41 (known as the K3 peptide) has been shown to form long, straight amyloid fibrils under acidic conditions either in isolation²⁹ or when disulfide bonded to its complementary 76–91 fragment,²³ the data suggest that this region is protected from aggregation by residual structure and/or the presence of a disulfide bond⁴¹ in the acid-denatured state of the full-length protein.

To further explore the hypothesis that non-random structure in the acid-unfolded state of β_2m influences fibril formation, we compared the apparent rate of

fibril elongation determined from seeded growth experiments with predictions derived using the Zygggregator algorithm modified to account for the low pH conditions²⁵ (Figs. 6a and 7a). Using this algorithm, the agreement between the predicted and observed aggregation rates was significant ($p < 0.001$) but not high ($R^2 = 0.53$) (Fig. 7a). For example, five variants (V27A, V37A, Y63A, Y66A and V82A) are predicted to have an identical Z_1^{agg} score (1.24), yet

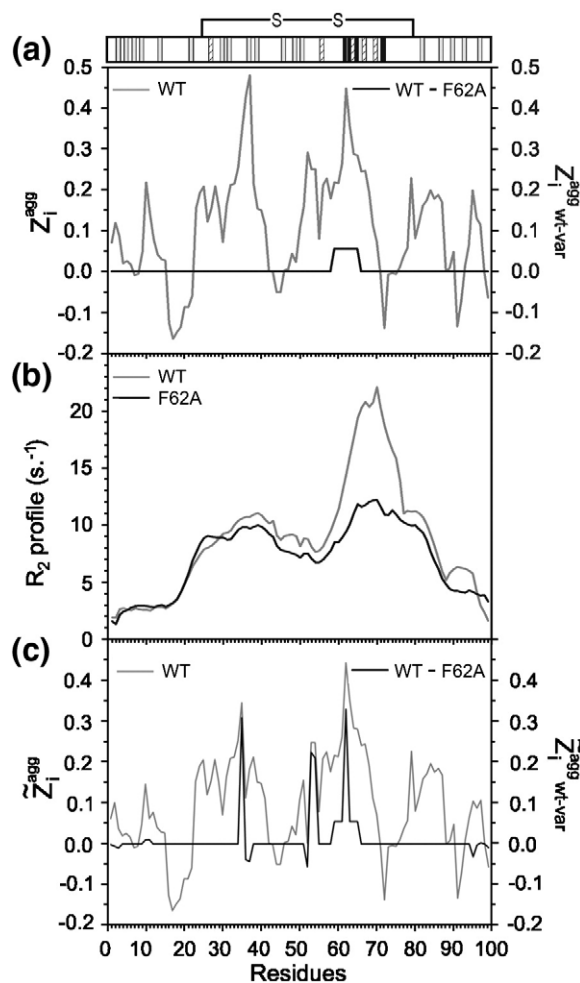


Fig. 6. Prediction of the aggregation properties of β_2m and an example variant, F62A. (a) Intrinsic propensity for aggregation of wild-type β_2m (grey, left axis) as predicted by the Zygggregator algorithm modified to acidic conditions.²⁵ The difference in aggregation propensity between wild-type β_2m and its variant F62A (black, right axis) is predicted by Z_1^{agg} to have only a small local effect on the intrinsic aggregation propensity of β_2m . A schematic illustration of the observed effects of amino acid substitution on the apparent rate of fibril elongation taken from Fig. 3d is shown above. The effects are classified as in Fig. 3, with rates not significantly altered compared with wild-type β_2m (grey), 2- to 4-fold change (hashed) and >4-fold change (black). (b) Comparison of the smoothed R_2 relaxation rates for wild-type β_2m and F62A β_2m . Plots of R_2 rates on a per-residue basis for all variants are shown in Fig. 8. (c) Comparison of the aggregation propensities of wild-type and the difference in aggregation propensities between wild-type and F62A β_2m obtained by modification of the Zygggregator algorithm by the measured R_2 values (\tilde{Z}_1^{agg}).

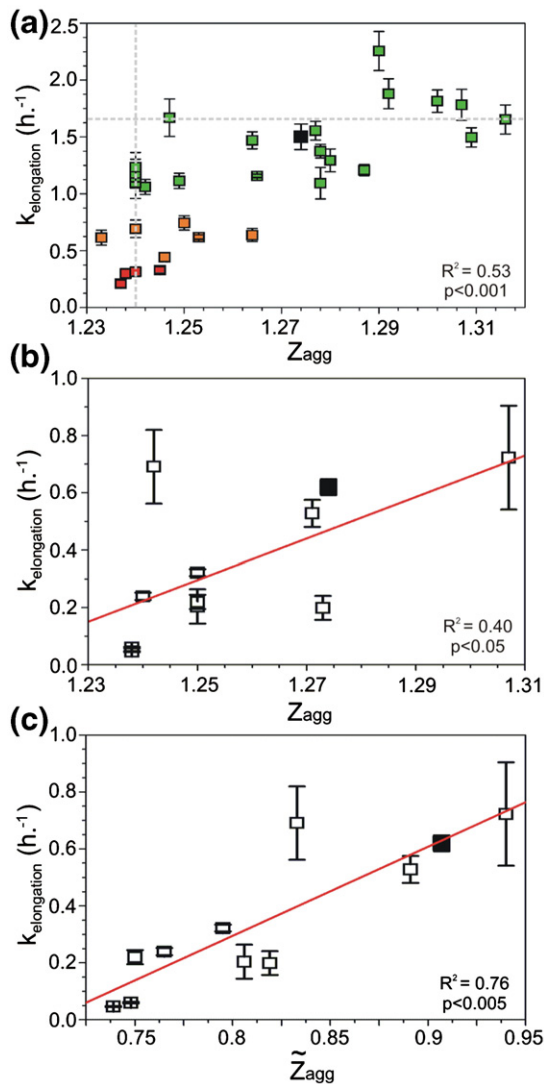


Fig. 7. Correlation of the predicted and observed rates of aggregation for $\beta_2\text{m}$ and its variants. (a) The observed elongation rates of wild-type $\beta_2\text{m}$ and its 30 variants (measured at 37 °C) compared with their predicted Z_{agg} scores calculated at low pH. The data are coloured according to Fig. 3d. (b) Comparison of the observed elongation rates of wild-type $\beta_2\text{m}$ and 10 variants (measured at 25 °C) with the predicted Z_{agg} scores at low pH; the red line indicates the correlation. (c) \tilde{Z}_{agg} scores (including the modulation of the Z_{agg} scores through the use of R_2 relaxation rates) of (b), the outlier where $k_{\text{elong}} = 0.69$ corresponds to F30A. The red line indicates the correlation. In all plots, wild-type $\beta_2\text{m}$ is represented as a black square.

they vary by 4-fold in their observed elongation rates (Fig. 7a, vertical grey line). Additionally, variants F22A and I46A have similar experimental rates of elongation ($\sim 1.68 \text{ h}^{-1}$), but are predicted to have Z_i^{agg} scores that differ substantially (1.25 and 1.32, respectively) (Fig. 7a, horizontal grey line). These data indicate that the prediction of intrinsic propensity for aggregation from the inherent properties of the constituent amino acids should be coupled to an assessment of the tendency to form intramolecular

interactions to accurately define the aggregation properties of $\beta_2\text{m}$ at pH 2.5.

As it has been observed by NMR measurements that acid-unfolded $\beta_2\text{m}$ contains significant non-random hydrophobic interactions,²¹ we hypothesised that these intramolecular interactions, in addition to the presence of the disulfide bond, could modulate the aggregation potential of the denatured chain. To account for this, we further generalised the Zyggregator algorithm to include the presence of intramolecular interactions, indicated by the magnitude of the R_2 values measured on a per-residue basis using NMR (Figs. 6b and 8) (Materials and Methods). The use of the R_2 relaxation data to account for the tendency to form intramolecular interactions in the low pH state of $\beta_2\text{m}$ is analogous to the use of protection factors from hydrogen exchange to rationalise the effects of the protection provided by the burial of the amyloidogenic regions in the native state described previously.²⁵ Modification of Z_i^{agg} by the R_2 value (termed \tilde{Z}_i^{agg}) also takes into account structural restraints dictated by the disulfide bond and allows the effect of specific amino acid substitutions on the intrinsic aggregation potential of the precursor state and its sensitivity to amino acid substitutions (Fig. 6c). For the analysis, we utilised a test set consisting of wild-type $\beta_2\text{m}$ and 10 variants (I7A, F30A, L40F, L40R, F62A, Y66A, Y66E, Y66S, Y67A and F70A) (Supplementary Table 2) for which both the R_2 values and the apparent elongation rates determined using seeded reactions (at 25 °C) were available (Fig. 8).¹⁹ Excluding the contribution of the NMR R_2 data (Fig. 7b), the correlation between the experimentally determined apparent elongation rates and those predicted using the low-pH, modified Zyggregator algorithm was not high ($R^2 = 0.40$) (Fig. 7b), indicating that the role of intramolecular interactions should be accounted for in order to perform accurate predictions. The incorporation of the R_2 measurements for each variant (Figs. 6c and 8) into the Zyggregator algorithm (see Materials and Methods) results in a significant improvement in the correlation with the corresponding experimental values ($p < 0.005$, $R^2 = 0.76$) (Fig. 7c), although some of the rates in the ~ 20 – 40 region are still overestimated (particularly for F30A), indicating that the R_2 measurements used here provide only an approximate estimation of the role of intramolecular interactions in modulating the intrinsic aggregation propensities. In the case of native states, we were able to show that the tendency to form intramolecular interactions could be predicted directly from the sequence, without the need of measuring protection factors from hydrogen exchange.¹¹ It could also be possible to estimate these tendencies in non-native states, thus avoiding the measurements of R_2 relaxation rates to perform the predictions of the overall aggregation propensities for different protein sequences. The results demonstrate that the residual structure in acid-denatured $\beta_2\text{m}$ is important in modulating the inherent aggregation propensity endowed by its amino acid sequence and that the ability of this structure to enhance or retard the observed aggregation rate is dependent upon the

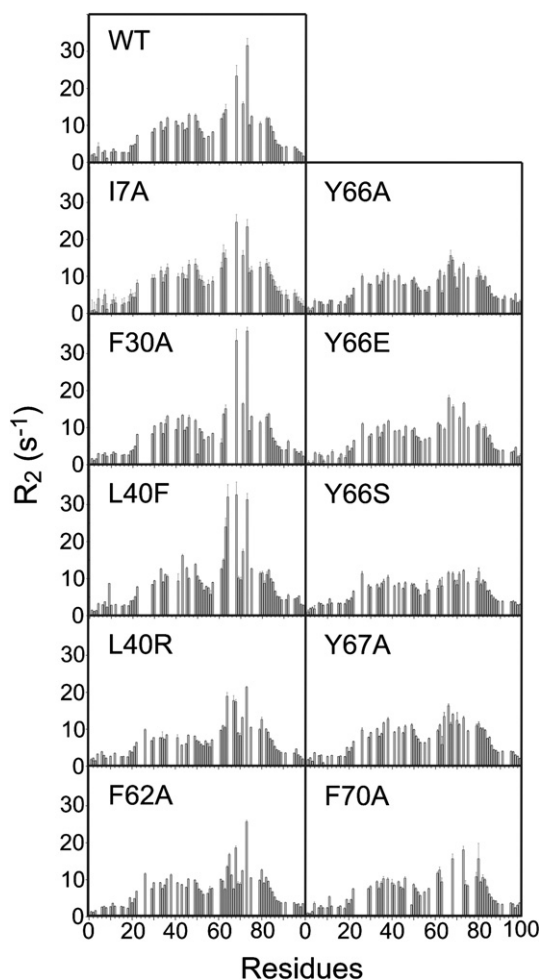


Fig. 8. Residue-specific R_2 rates of wild-type β_2m and 10 variants at pH 2.5 and 25 °C. R_2 rates were measured at 500 MHz in water as described in Materials and Methods.

location of intramolecular interactions in the polypeptide chain. They also show how the inclusion of information describing these structural characteristics permits accurate prediction of the effect of even subtle mutations on the rate of fibril formation.

Discussion

The results presented show that only one of the intrinsically aggregation-prone regions of β_2m predicted using algorithms or peptide studies defines the aggregation potential of the complete polypeptide chain under the denaturing conditions that we have examined. This finding indicates that intramolecular and intermolecular interactions compete against each other to determine the overall aggregation process even in denatured proteins, so that the intrinsic propensities for aggregation and for forming soluble monomeric species modulate each other. Indeed, previous studies have implicated long-range intramolecular contacts involving the NAC region of α -synuclein in defining its aggregation properties,^{42,43} whilst the presence of one particular hydrophobic

cluster in denatured lysozyme at pH 2 has been shown to play a major role in preventing fibril formation.⁴⁴ As we have demonstrated here, incorporation of the competing influences on collapse, folding and aggregation (Fig. 1) into prediction algorithms enables regions critical for aggregation to be identified in the context of full-length polypeptide sequences and the effects of subtle mutations to be quantified. Thus, whilst the intrinsic aggregation propensity can provide information as to which regions of the sequence have a potential to aggregate, any consideration of fibril formation rates must be careful to take the structural properties of monomeric and oligomeric species into account. Taken together, these insights provide a deeper understanding of the fundamental forces that control the competition between folding, misfolding and aggregation. In addition, they should facilitate the development of rational design strategies to identify reagents capable of inhibiting aggregation by targeting regions known to be critical for the aggregation process.

Materials and Methods

The *Escherichia coli* strain BL21(DE3) pLysS was obtained from Promega. Q-Sepharose and all other reagents were purchased from the Sigma-Aldrich Chemical Company. Spectrapore dialysis membrane (molecular mass cutoff, 3500 Da) was acquired from Spectrum Laboratories, Inc. Superdex 75 was purchased from Amersham Biosciences. Carbenicillin was obtained from Melford Chemicals. Deuterated solvents were purchased from Fluorochem Ltd., and oligonucleotides were from MWG Biotech.

Generation of variants and protein purification

Single amino acid substitutions and the variant, $\Delta N6$, were generated within the template gene pINK as previously described.^{19,45} $\Delta C83$ was created by replacement of the codons for His84 and His85 with stop codons. Unlabelled and ¹⁵N-labelled proteins were expressed and purified to $\geq 95\%$ purity as previously described.¹⁹

Kinetic analysis of fibril formation

Data were recorded as previously described¹⁹ with small variations in the methodology. Briefly, stock solutions of protein in ~ 5 mM HCl were centrifuged at 543,000g for 1 h (20 °C) and only the upper two-thirds of the solution was used for fibril growth experiments. For spontaneous fibril growth reactions, 84 μM β_2m ($n=6$ and $n=9$ for wild-type and variant β_2m samples, respectively) was incubated in pH 2.5 buffer (25 mM sodium phosphate, 25 mM sodium acetate and 0.02% azide) containing 10 μM ThT with agitation (600 rpm) in the wells of 96-well black-wall plates (Costar) and sealed with clear sealing film (Axygen). Measurements were recorded at 25 or 37 °C, as reported. ThT fluorescence was recorded at 5-min intervals using a FLUOstar Optima reader (BMG) as previously described.¹⁹ Comparison of simultaneous ThT and intrinsic tryptophan fluorescence measurements showed that ThT did not interfere with fibril formation, and its fluorescence is a reliable probe of fibril formation (data not shown).²⁰ $\Delta C83$ formed fibrils

after incubation of the same samples at 200 rpm for >4 weeks at 37 °C.

In the case of seeded reactions at 37 °C, 5% (v/v) seed was added to each reaction and fibril growth was monitored as described above. The seed consisted of wild-type β_2m fibrils, formed in pH 2.5 buffer, agitated (200 rpm) for 1 week at 37 °C and subsequently fragmented by three freeze–thaw cycles. The seeds (Fig. 4b) were stored at –80 °C, and the same batch of seeds was used for all reactions. For comparison with NMR measurements, the apparent rates of elongation determined using seeded reactions of the second test set of variants were measured at 25 °C with the addition of 2% (v/v) wild-type seed.¹⁹

All data were normalised to the final ThT signal (after 72 h growth), and the resulting curves were used to determine the lag times and apparent rates of elongation. The lag time was obtained by fitting a straight line to the slope of the growth phase between approximately 30% and 70% of the maximum amplitude, and the time at which this line intersected the baseline was taken as the lag time (Fig. 3b, inset). The apparent elongation rates (k) were determined by fitting a tangent to the initial part of the growth curve (Fig. 3d, inset).

Electron microscopy

Samples were examined either undiluted or with the use of a 1:10 dilution in 5 mM HCl on freshly ionised formvar- and carbon-coated electron microscopy grids (Agar). The grids were rinsed with 5 mM HCl and stained with 4% (w/v) uranyl acetate. All images were taken using a Philips CM10 electron microscope operating at 80 keV.

NMR spectroscopy

All NMR experiments were carried out using 5 mg ml⁻¹ ¹⁵N-labelled β_2m in 90% (v/v) H₂O/10% (v/v) ²H₂O at pH 2.5 and 25 °C. Experiments were performed using a Varian Unity Inova spectrometer operating at a proton frequency of 500 MHz. Gradient-enhanced ¹H-¹⁵N heteronuclear single quantum coherence spectra were acquired using 128 complex points and 16 scans per increment with spectral widths of 4508 and 1200 Hz in the ¹H and ¹⁵N dimensions, respectively. Watergate solvent suppression was used, and all NMR data were processed using NMRPipe.⁴⁶ The data were apodised using a cosine bell function, followed by zero filling and Fourier transform. The 2D spectra were analysed in NMRView.⁴⁷ Backbone ¹⁵N transverse R_2 relaxation measurements were carried out using a series of 11 experiments with mixing times ranging from 16.32 to 456.96 ms, as previously described.^{19,48} Duplicate points and the spectral noise levels were used to obtain an estimate of the error.

Calculation of the aggregation propensities

The aggregation propensities at acidic pH were calculated using Z_{agg} scores as described by Tartaglia *et al.*,¹¹ but using pH-dependent propensity scales.²⁵ In our previous work, the intrinsic aggregation propensities due to the presence of residual structure were modulated through the use of hydrogen exchange profiles.¹¹

$$\bar{Z}_i^{agg} = Z_i^{agg}(1 - \epsilon HX_i)$$

where $\epsilon = 1/15$ and HX_i is the protection factor of residue i .¹¹

In analogy with this formula, we used here the R_2 relaxation rates for modulating the aggregation propensities. The R_2 data were smoothed using a seven-residue window and labelled as ρ_i . The Z_i^{agg} profiles were modified by taking into account the R_2 data using the following phenomenological formula:

$$\tilde{Z}_i^{agg} = \begin{cases} (Z_i^{agg} - \alpha\rho_i) * \left(1 + \alpha \frac{\rho_i}{\rho_{max}}\right) & \text{if } Z_i^{agg} > 0 \\ Z_i^{agg} & \text{otherwise} \end{cases} \quad (1)$$

where the constant $\alpha = 1.32$ and $\rho_{max} = 30$ represents the maximum of ρ_i .

The analytical form of the equation captures the fact that regions such as 20–40 and 50–70 are involved in both aggregation and folding. As indicated in the term $Z_i^{agg} - \alpha\rho_i$, the aggregation propensity decreases as a consequence of the propensity to form native contacts, whilst the multiplicative term $\left(1 + \alpha \frac{\rho_i}{\rho_{max}}\right)$ introduces the effect that an intrinsic propensity to form intramolecular contacts comes at the cost of an increased propensity to form intermolecular contacts. Here, Eq. (1) represents the expansion of our previous formula¹¹ that allows the calculation of the aggregation propensity using the structural information based on the conformation properties of the precursor state.

The global score is calculated as follows:²⁵

$$\tilde{Z}_{agg} = \frac{\sum_{i=1}^{length} \tilde{Z}_i^{agg} * \theta(\tilde{Z}_i^{agg})}{\left(\sum_{i=1}^{length} \theta(\tilde{Z}_i^{agg})\right)^\gamma} L^\beta$$

where $\beta = 0.44$ and $\gamma = 0.75$ and L is the length of the amino acid sequence.

The function $\theta(\tilde{Z}_i^{agg}) = 1$ if $\tilde{Z}_i^{agg} > 0$ and $\theta(\tilde{Z}_i^{agg}) = 0$ otherwise.

Acknowledgements

We thank S. Jones and T. Jahn for cloning some of the variant proteins. We also thank T. Eichner and B. Tynan-Connelly for expressing and purifying the P to G variants. K.E.R. and G.W.P. were supported by the Wellcome Trust. G.G.T. and M.V. were supported by the Leverhulme Trust, and M.V. was supported by the European Molecular Biology Organization and the Royal Society. We acknowledge the members of the S. E.R. and M.V. groups for many valuable discussions.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.jmb.2009.04.042](https://doi.org/10.1016/j.jmb.2009.04.042)

References

1. Westermarck, P., Benson, M. D., Buxbaum, J. N., Cohen, A. S., Frangione, B., Ikeda, S. *et al.* (2007). A primer of amyloid nomenclature. *Amyloid*, **14**, 179–183.

2. Sunde, M. & Blake, C. (1997). The structure of amyloid fibrils by electron microscopy and X-ray diffraction. *Adv. Protein Chem.* **50**, 123–159.
3. Chiti, F. & Dobson, C. M. (2006). Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333–366.
4. Fowler, D. M., Koulov, A. V., Balch, W. E. & Kelly, J. W. (2007). Functional amyloid—from bacteria to humans. *Trends Biochem. Sci.* **32**, 217–224.
5. Cherny, I. & Gazit, E. (2008). Amyloids: not only pathological agents but also ordered nanomaterials. *Angew. Chem., Int. Ed. Engl.* **47**, 4062–4069.
6. Selkoe, D. J. (2003). Folding proteins in fatal ways. *Nature*, **426**, 900–904.
7. Sawaya, M. R., Sambashivan, S., Nelson, R., Ivanova, M. I., Sievers, S. A., Apostol, M. I. *et al.* (2007). Atomic structures of amyloid cross- β spines reveal varied steric zippers. *Nature*, **447**, 453–457.
8. Wasmer, C., Lange, A., Van Melckebeke, H., Siemer, A. B., Riek, R. & Meier, B. H. (2008). Amyloid fibrils of the HET-s(218–289) prion form a β -solenoid with a triangular hydrophobic core. *Science*, **319**, 1523–1526.
9. Margittai, M. & Langen, R. (2008). Fibrils with parallel in-register structure constitute a major class of amyloid fibrils: molecular insights from electron paramagnetic resonance spectroscopy. *Q. Rev. Biophys.* **41**, 265–297.
10. Tycko, R. (2006). Molecular structure of amyloid fibrils: insights from solid-state NMR. *Q. Rev. Biophys.* **39**, 1–55.
11. Tartaglia, G. G., Pawar, A. P., Campioni, S., Dobson, C. M., Chiti, F. & Vendruscolo, M. (2008). Prediction of aggregation-prone regions in structured proteins. *J. Mol. Biol.* **380**, 425–436.
12. Sekijima, Y., Kelly, J. W. & Ikeda, S. (2008). Pathogenesis of and therapeutic strategies to ameliorate the trans-thyretin amyloidosis. *Curr. Pharm. Des.* **14**, 3219–3230.
13. Ricagno, S., Raimondi, S., Giorgetti, S., Bellotti, V. & Bolognesi, M. (2009). Human β_2 -microglobulin W60V mutant structure: implications for stability and amyloid aggregation. *Biochem. Biophys. Res. Commun.* **380**, 543–547.
14. Dumoulin, M., Canet, D., Last, A. M., Pardon, E., Archer, D. B., Muyldermans, S. *et al.* (2005). Reduced global cooperativity is a common feature underlying the amyloidogenicity of pathogenic lysozyme mutations. *J. Mol. Biol.* **346**, 773–788.
15. Guo, Z. & Eisenberg, D. (2007). The mechanism of the amyloidogenic conversion of T7 endonuclease I. *J. Biol. Chem.* **282**, 14968–14974.
16. Quintas, A., Vaz, D. C., Cardoso, I., Saraiva, M. J. & Brito, R. M. (2001). Tetramer dissociation and monomer partial unfolding precedes protofibril formation in amyloidogenic transthyretin variants. *J. Biol. Chem.* **276**, 27207–27213.
17. Furukawa, Y., Kaneko, K., Yamanaka, K., O'Halloran, T. V. & Nukina, N. (2008). Complete loss of post-translational modifications triggers fibrillar aggregation of SOD1 in the familial form of amyotrophic lateral sclerosis. *J. Biol. Chem.* **283**, 24167–24176.
18. Page, L. J., Suk, J. Y., Huff, M. E., Lim, H. J., Venable, J., Yates, J. *et al.* (2005). Metalloendoprotease cleavage triggers gelsolin amyloidogenesis. *EMBO J.* **24**, 4124–4132.
19. Platt, G. W., Routledge, K. E., Homans, S. W. & Radford, S. E. (2008). Fibril growth kinetics reveal a region of β_2 -microglobulin important for nucleation and elongation of aggregation. *J. Mol. Biol.* **378**, 251–263.
20. Xue, W. F., Homans, S. W. & Radford, S. E. (2008). Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly. *Proc. Natl Acad. Sci. USA*, **105**, 8926–8931.
21. Platt, G. W., McParland, V. J., Kalverda, A. P., Homans, S. W. & Radford, S. E. (2005). Dynamics in the unfolded state of β_2 -microglobulin studied by NMR. *J. Mol. Biol.* **346**, 279–294.
22. Jones, S., Manning, J., Kad, N. M. & Radford, S. E. (2003). Amyloid-forming peptides from β_2 -microglobulin—insights into the mechanism of fibril formation *in vitro*. *J. Mol. Biol.* **325**, 249–257.
23. Hasegawa, K., Ohhashi, Y., Yamaguchi, I., Takahashi, N., Tsutsumi, S., Goto, Y. *et al.* (2003). Amyloidogenic synthetic peptides of β_2 -microglobulin—a role of the disulfide bond. *Biochem. Biophys. Res. Commun.* **304**, 101–106.
24. Ivanova, M. I., Thompson, M. J. & Eisenberg, D. (2006). A systematic screen of β_2 -microglobulin and insulin for amyloid-like segments. *Proc. Natl Acad. Sci. USA*, **103**, 4079–4082.
25. Tartaglia, G. G. & Vendruscolo, M. (2008). The Zyggregator method for predicting protein aggregation propensities. *Chem. Soc. Rev.* **37**, 1395–1401.
26. 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. *Nat. Biotechnol.* **22**, 1302–1306.
27. Conchillo-Sole, O., de Groot, N. S., Aviles, F. X., Vendrell, J., Daura, X. & Ventura, S. (2007). AGGRES-CAN: a server for the prediction and evaluation of “hot spots” of aggregation in polypeptides. *BMC Bioinformatics*, **8**, 65.
28. Uversky, V. N. (2008). Amyloidogenesis of natively unfolded proteins. *Curr. Alzheimer Res.* **5**, 260–287.
29. Kozhukh, G. V., Hagihara, Y., Kawakami, T., Hasegawa, K., Naiki, H. & Goto, Y. (2002). Investigation of a peptide responsible for amyloid fibril formation of β_2 -microglobulin by *achromobacter* protease I. *J. Biol. Chem.* **277**, 1310–1315.
30. Ivanova, M. I., Gingery, M., Whitson, L. J. & Eisenberg, D. (2003). Role of the C-terminal 28 residues of β_2 -microglobulin in amyloid fibril formation. *Biochemistry*, **42**, 13536–13540.
31. Ivanova, M. I., Sawaya, M. R., Gingery, M., Attinger, A. & Eisenberg, D. (2004). An amyloid-forming segment of β_2 -microglobulin suggests a molecular model for the fibril. *Proc. Natl Acad. Sci. USA*, **101**, 10584–10589.
32. Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L. & Wiley, D. C. (1987). Structure of the human class I histocompatibility antigen, HLA-A2. *Nature*, **329**, 506–512.
33. Smith, D. P., Jones, S., Serpell, L. C., Sunde, M. & Radford, S. E. (2003). A systematic investigation into the effect of protein destabilisation on β_2 -microglobulin amyloid formation. *J. Mol. Biol.* **330**, 943–954.
34. Smith, D. P. & Radford, S. E. (2008). Hemodialysis related amyloidosis. In *Protein Misfolding Diseases: Current and Emerging Principles and Possible Therapies* (Ramirez-Alvarado, M., Kelly, J. & Dobson, C. M., eds). John Wiley & Sons, Inc., in press.
35. Sanchez de Groot, N., Pallares, I., Aviles, F. X., Vendrell, J. & Ventura, S. (2005). Prediction of “hot spots” of aggregation in disease-linked polypeptides. *BMC Struct. Biol.* **5**, 18.
36. Esposito, G., Michelutti, R., Verdonesi, G., Viglino, P.,

- Hernandez, H., Robinson, C. V. *et al.* (2000). Removal of the N-terminal hexapeptide from human β_2 -microglobulin facilitates protein aggregation and fibril formation. *Protein Sci.* **9**, 831–845.
37. Vendruscolo, M., Paci, E., Dobson, C. M. & Karplus, M. (2001). Three key residues form a critical contact network in a protein folding transition state. *Nature*, **409**, 641–645.
38. 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.
39. Smith, D. P. & Radford, S. E. (2001). Role of the single disulphide bond of β_2 -microglobulin in amyloidosis *in vitro*. *Protein Sci.* **10**, 1775–1784.
40. Katou, H., Kanno, T., Hoshino, M., Hagihara, Y., Tanaka, H., Kawai, T. *et al.* (2002). The role of disulfide bond in the amyloidogenic state of β_2 -microglobulin studied by heteronuclear NMR. *Protein Sci.* **11**, 2218–2229.
41. Hamada, D., Tanaka, T., Tartaglia, G. G., Pawar, A. P., Vendruscolo, M., Kawamura, M. *et al.* (2009). Competition between folding, native-state dimerisation and amyloid aggregation in β -lactoglobulin. *J. Mol. Biol.* **386**, 878–890.
42. Wu, K. P., Kim, S., Fela, D. A. & Baum, J. (2008). Characterization of conformational and dynamic properties of natively unfolded human and mouse α -synuclein ensembles by NMR: implication for aggregation. *J. Mol. Biol.* **378**, 1104–1115.
43. Ehrnhoefer, D. E., Bieschke, J., Boeddrich, A., Herbst, M., Masino, L., Lurz, R. *et al.* (2008). EGCG redirects amyloidogenic polypeptides into unstructured, off-pathway oligomers. *Nat. Struct. Mol. Biol.* **15**, 558–566.
44. Mishima, T., Ohkuri, T., Monji, A., Imoto, T. & Ueda, T. (2007). A particular hydrophobic cluster in the residual structure of reduced lysozyme drastically affects the amyloid fibrils formation. *Biochem. Biophys. Res. Commun.* **356**, 769–772.
45. Myers, S. L., Jones, S., Jahn, T. R., Morten, I. J., Tennent, G. A., Hewitt, E. W. *et al.* (2006). A systematic study of the effect of physiological factors on β_2 -microglobulin amyloid formation at neutral pH. *Biochemistry*, **45**, 2311–2321.
46. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
47. Johnson, B. & Blevins, R. (1994). NMRView: a computer programme for the visualization and analysis of NMR data. *J. Biomol. NMR*, **4**, 603–614.
48. Farrow, N. A., Zhang, O., Forman-Kay, J. D. & Kay, L. E. (1994). A heteronuclear correlation experiment for simultaneous determination of ^{15}N longitudinal decay and chemical exchange rates of systems in slow equilibrium. *J. Biomol. NMR*, **4**, 727–734.