doi:10.1016/j.jmb.2008.05.013 J. Mol. Biol. (2008) 380, 425–436

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Prediction of Aggregation-Prone Regions in Structured Proteins

Gian Gaetano Tartaglia[1], Amol P. Pawar[1], Silvia Campioni[2], Christopher M. Dobson[1], Fabrizio Chiti[2] and Michele Vendruscolo[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*

Received 9 January 2008; received in revised form 2 May 2008; accepted 8 May 2008
Available online 13 May 2008

Edited by J. Weissman

We present a method for predicting the regions of the sequences of peptides and proteins that are most important in promoting their aggregation and amyloid formation. The method extends previous approaches by allowing such predictions to be carried out for conditions under which the molecules concerned can be folded or contain a significant degree of persistent structure. In order to achieve this result, the method uses only knowledge of the sequence of amino acids to estimate simultaneously both the propensity for folding and aggregation and the way in which these two types of propensity compete. We illustrate the approach by its application to a set of peptides and proteins both associated and not associated with disease. Our results show not only that the regions of a protein with a high intrinsic aggregation propensity can be identified in a robust manner but also that the structural context of such regions in the monomeric form is crucial for determining their actual role in the aggregation process.

Keywords: Aggregation; Misfolding; Prion disease; Alzheimer's disease; Parkinson's disease

Introduction

It is becoming increasingly clear that specific regions of the amino acid sequences of polypeptide chains, sometimes known as "aggregation-prone" regions,[1] play a major role in determining their tendency to aggregate and, ultimately, to form organized structures such as amyloid fibrils.[1–3] Strong support for this view has been provided by analyzing the effects of mutations on the aggregation propensities of specific peptides and proteins,[4] and through the determination of high-resolution structural models illustrating that specific segments of polypeptide chains constitute the highly ordered cores of the fibrils.[5–10] The existence of aggregation-prone regions has suggested the way in which rational mutagenesis can reduce the problem of aggregation in biotechnology and in the development of new drugs.[11,12] In addition, it has suggested therapeutic strategies that specifically target these regions in order to reduce their tendency to promote the formation of ordered intermolecular assemblies.[13–15]

The main physicochemical factors that promote the aggregation of unfolded polypeptide chains have recently been characterized;[16,17] based on these findings, several algorithms have been proposed to predict "aggregation propensity profiles" that enable the identification of regions with a high intrinsic propensity for aggregation.[1,3,18–24] We have previously shown the strength of this approach in predicting the aggregation-prone regions of polypeptide chains that are largely unstructured under physiological conditions, including the amyloid β (Aβ) peptide associated with Alzheimer's disease and α-synuclein, a natively unfolded protein whose aggregation has been linked to Parkinson's disease.[1]

In this article, we generalize this approach to predict the regions that promote the aggregation of structured globular proteins or proteins that contain both structured and unstructured domains. In such an enterprise, it is essential to consider the possibility that regions with a high intrinsic propensity for aggregation may be buried inside stable and often highly cooperative structural elements and, therefore, may be unable in such states to form the specific intermolecular interactions that lead to aggregation. Shielded in this way, they may become unable to play a major role in the aggregation process, although,

*Corresponding author. E-mail address: mv245@cam.ac.uk.

Abbreviations used: Aβ, amyloid β; IAPP, islet amyloid polypeptide; hIAPP, human IAPP; EPR, electron paramagnetic resonance; TTR, transthyretin; hPrP, human prion protein; $hPrP^{C}$, cellular isoform of the prion protein; $hPrP^{Sc}$, β-sheet-rich aggregated form of the prion protein.

following mutations that destabilize the native structure, they might acquire this ability. In order to be able to take into consideration the tendency of a given region of a protein sequence to adopt a folded conformation, we exploit the possibility of predicting the local stability of the various regions of a protein from knowledge of its sequence.[25] In essence, given the amino acid sequence of a protein, we show here how it is possible to combine the predictions of the intrinsic aggregation propensity profiles with those for folding into stable structures to determine new aggregation propensity profiles of structured or partially structured proteins that account for the influence of the structural context. We illustrate this approach through its application to the prediction of aggregation profiles for a range of peptides and proteins whose aggregation propensities have been characterized experimentally in particular detail.

Results

Experimentally, aggregation-prone regions have been identified by a range of different techniques, including mutational analysis of the kinetics of the aggregation process[4,7] or of the stability of the amyloid fibrils,[26,27] structural analysis of the cores of amyloid fibrils,[10] fluorescence techniques,[7] and the study of the aggregation of peptide fragments extracted from wild-type proteins.[28,29] These probes report on different aspects of the dynamics of the aggregation process and of the thermodynamics of the amyloid state. Since the predictions that we perform are based on the analysis of mutational effects on the kinetics of aggregation,[1,16,17] we are interested both in assessing the quality of the predictions of the regions that are most important in promoting the aggregation process, and in exploring the relationship between such aspect and the other factors that may affect the formation and stability of globular forms and amyloid fibrils.

Prediction of the aggregation propensities of unstructured polypeptides

We first present predictions of the aggregation propensity profiles of five unstructured human peptides of less than 50 residues and of one natively unfolded protein that are involved in protein deposition diseases: $A\beta_{1-42}$, calcitonin, glucagon, the second WW domain of CA150, islet amyloid polypeptide (IAPP), and α-synuclein (Fig. 1). In addition to the intrinsic aggregation propensity profile Z_i^{agg}, which is calculated with a procedure similar to the one that we have already described[1] (see Methods), we present a second type of profile $\tilde{Z}_i^{agg}$, which takes into account the propensities of the different regions of a polypeptide chain to form stable folded structures (see Methods).

Aβ1–42

The Aβ peptide is the main constituent of the extracellular deposits characteristic of Alzheimer's disease.[39] This peptide is found in the human brain predominantly in two forms of: 40 and 42 amino acids in length ($A\beta_{1-40}$ and $A\beta_{1-42}$, respectively). The intrinsic aggregation profile Z_i^{agg} of $A\beta_{1-42}$ (red line in Fig. 1a) reveals two regions of high aggregation propensity (those above the $Z_i^{agg}=1$ threshold): the central region (residues 18–22) and the C-terminal region (residues 32–42). Both these regions play an important structural role in the current models of the structures of the $A\beta_{1-40}$[8,40] and $A\beta_{1-42}$[30] peptides in their amyloid forms. The recalculated aggregation propensity profile $\tilde{Z}_i^{agg}$ (black line in Fig. 1a), which takes into account the tendency of the monomeric form of $A\beta_{1-42}$ to adopt a persistent conformation in solution, does not reveal remarkable changes relative to Z_i^{agg}, in agreement with the well-established notion that the peptide is largely unstructured in solution with no persistently formed folded structure.

Calcitonin

Human calcitonin is a 32-residue polypeptide hormone involved in calcium regulation and bone dynamics that has been shown to be present as amyloid fibrils in patients with medullary carcinoma of the thyroid.[41] In addition, fibrils can also form in samples prepared *in vitro* and designed for therapeutic use, and represent a considerable limitation on its administration to patients.[42] By calculating the aggregation profile $\tilde{Z}_i^{agg}$, we predict a high aggregation potential for the 12-residue N-terminal region and for residues 18–19 and 27–28 (Fig. 1b). Although some information on the role of the 15–19 and C-terminal regions in the aggregation of calcitonin has been published,[29,30,32,39–42] an exhaustive experimental study aimed at identifying the regions of the sequence that promote aggregation of the full-length hormone has not been yet reported, making it difficult to compare our profiles with experimental data. We did not predict an intrinsic tendency for the monomeric form of this short peptide to form a persistent structure, which is consistent with available experimental evidence.[43] The intrinsic aggregation propensity profile Z_i^{agg} is therefore close to the $\tilde{Z}_i^{agg}$ profile.

Glucagon

Glucagon is a 29-residue hormone that participates in carbohydrate metabolism. Since it promotes the release of glucose from the liver to the blood, this polypeptide hormone has been used in the treatment of hypoglycemia.[44] Glucagon has been shown readily to form amyloid fibrils under acidic conditions.[33] The regions encompassing residues 6–10 and 23–27 appear to be important for fibril formation, while the central region (residues 13–18 and 22) plays a major role in determining the morphology of the fibrils themselves.[33] In agreement with these experimental findings, we predict the N-terminal region (in particular, residues 4–8) and the C-terminal region (in particular, residues 24–27) to be highly aggregation-prone (Fig. 1c). As in the cases of $A\beta_{1-42}$ and calcitonin, glucagon is not highly structured in its

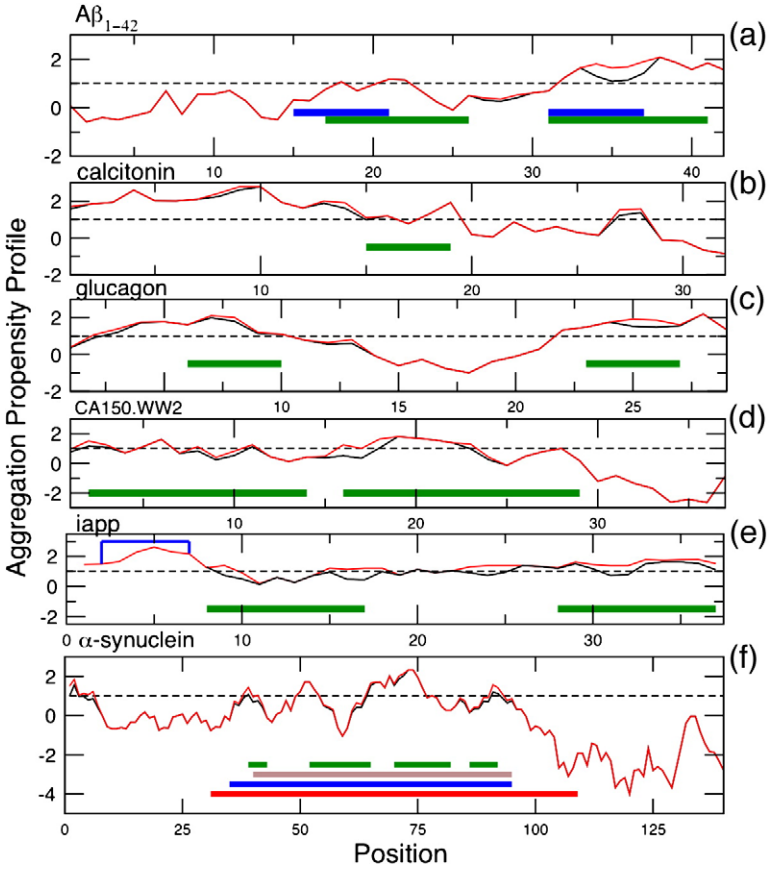

Fig. 1. Aggregation propensity profiles of six unstructured human polypeptide chains involved in protein deposition diseases. Red lines indicate the intrinsic aggregation propensity profiles Z_i^{agg}. Black lines indicate the aggregation propensities calculated by taking into account the structural protection provided by the folded form of the protein $\tilde{Z}_i^{agg}$. (a) $A\beta_{1-42}$: the green and blue horizontal bars indicate regions of the sequence found to form the core of the fibrils as determined with solid-state NMR measurement[30] and site-directed spin labeling coupled with EPR,[31] respectively. (b) Calcitonin: the green horizontal bar indicates the region of residues 15–19 found to form the core of the fibrils.[29,32] (c) Glucagon: the green horizontal bars indicate the regions of the sequence found to promote aggregation of this peptide using a protein engineering approach.[33] (d) The second WW domain of human CA150: the green horizontal bars indicate the segments that form the cross-β-core in a structure recently proposed from solid-state NMR measurements.[34] (e) IAPP: the green bars show the regions that have been characterized as β-strands in the amyloid fibrils by using solid-state NMR spectroscopy;[35] the blue line indicates the position of the disulfide bridge between residues 2 and 7. (f) α-Synuclein: the blue, gray, and red horizontal bars indicate the region of the sequence that appears to be structured in the fibrils from site-directed spin labeling coupled with EPR,[36] hydrogen/deuterium exchange,[37] and limited proteolysis.[38] The regions adopting a β-strand conformation within such region, as revealed by solid-state NMR measurements, are indicated by green horizontal bars; due to experimental uncertainties, the boundaries of such strands are only approximates.[5]

monomeric form; consistent with the results of the intrinsic aggregation propensity, the Z_i^{agg} profile is close to the $\tilde{Z}_i^{agg}$ profile (Fig. 1c).

CA150.WW2

The second WW domain of human CA150, a protein that is codeposited with huntingtin in Huntington's disease, is a 40-residue protein that has been shown to form amyloid fibrils *in vitro* under physiological conditions.[45] The structure of this WW domain in the amyloid protofilament was recently characterized by solid-state NMR spectroscopy,[34] indicating that residues 2–14 and 16–29 constitute the core of the fibrils. These experimental results are in agreement with those calculated here (Fig. 1d). Again, no significant changes in the aggregation propensity profile were observed when considering the propensity to fold (Fig. 1d).

IAPP

Assemblies of the 37-residue human IAPP (hIAPP) represent the major component of the pancreatic amyloid deposits associated with type II diabetes.[46] Solid-state NMR studies of the structures of the amyloid fibrils formed by hIAPP have revealed that the regions of residues 8–17 and 28–37[35] (green bars in Fig. 1e) form β-strands in the amyloid state. The regions that form the structural core of the hIAPP amyloid fibril, however, need not be identical with those that promote amyloid formation,[35] and there is evidence that peptides corresponding to the regions of residues 20–25, 24–29, 22–27,[47,48] and 22–28 with the F23Y mutation[49] are important in modulating the aggregation process. We predicted only a marginal intrinsic tendency for the monomeric form of this short peptide to form a persistent structure; therefore, the intrinsic aggregation propensity profile Z_i^{agg} is rather close to the $\tilde{Z}_i^{agg}$ profile. Both profiles do not exhibit regions of negative aggregation propensities, indicating that hIAPP's overall tendency to aggregate is quite high. Our results also suggest that the disulfide bond between residues 2 and 7 has the likely effect of limiting the high aggregation propensity of the N-terminal region of hIAPP.

α-Synuclein

Human α-synuclein is known to self-assemble into intracellular inclusions in dopaminergic neurons of patients suffering from Parkinson's disease. Using an

array of experimental techniques, including solid-state NMR, limited proteolysis, hydrogen/deuterium exchange, and site-directed spin labeling/electron paramagnetic resonance (EPR), it was found that the central region (approximately residues 30–95) of this normally natively unfolded protein forms the core of the fibrils.[5,36–38,50] All the four peaks identified in both aggregation propensity profiles Z_i^{agg} and $\tilde{Z}_i^{agg}$ are located within this central region of the sequence (Fig. 1f). Moreover, the four peaks appear to correspond to the regions found to form the β-core of the fibrils using solid-state NMR measurements.[5]

Prediction of the aggregation profiles of globular proteins

The approach presented here is specifically designed to include the prediction of those regions of the amino acid sequence of a globular protein that promote its ordered aggregation starting from a structured state. In such cases, it is normally necessary to destabilize the structure to enhance the accessibility of the polypeptide main chain and hydrophobic side chains in order for the aggregation process to occur. In this section, we discuss four human proteins that have been shown to aggregate under such conditions.

Lysozyme

Human lysozyme has been found to form amyloid fibrils *in vivo* in patients suffering from hereditary amyloidosis carrying nonconservative point mutations in the lysozyme gene.[50–52] The aggregation propensity profile of wild-type human lysozyme $\tilde{Z}_i^{agg}$, calculated by taking into account the structural protection in the native state, as predicted from the sequence, does not exhibit any region above the $\tilde{Z}_i^{agg} = 1$ threshold (Fig. 2a, black line). This result is consistent with the observations that lysozyme must be destabilized *in vitro* in order for it to aggregate and that amyloid disease is only found as a result of destabilizing familiar mutations.[53] By calculating the intrinsic aggregation propensity profile Z_i^{agg} for wild-type human lysozyme, we identified six aggregation-prone regions (residues 24–32, 37–46, 53–67, 75–81, 92–93, and 124–127; i.e., those with a Z_i^{agg} score of ≥ 1) (Fig. 2a). These predictions are of particular interest in the light of recent experimental observations that the regions of the sequence that include residues 26–123 and 32–108 (Fig. 2a, green bars) are highly resistant to proteolysis once converted into the amyloid state.[53]

In order to clarify the relationship between the tendency to remain folded or the tendency to aggregate, we compared the structural protection and the aggregation propensity at a residue-specific level. The aggregation propensity was measured by the Z_i^{agg} score, and the structural protection was measured by the $\ln P_i$ score[25] (see Methods), providing a prediction of the local stability of the region comprising a particular residue (Fig. 3a). The calculation of the $\ln P_i$ score either from the sequence or from the native structure results in very similar $\tilde{Z}_i^{agg}$ profiles (blue and green lines, respectively, in Fig. 2a). In the plot of $\ln P_i$ *versus* Z_i^{agg}, regions of high aggregation propensity and low structural stability in the folded state, which are most likely to play an important role in the first stages of the aggregation process, are found on the right lower quadrant of the plot. We predict residues 25 and 78–80 (labeled in bold in Fig. 3a) to have the highest aggregation propensity and the lowest structural protection.

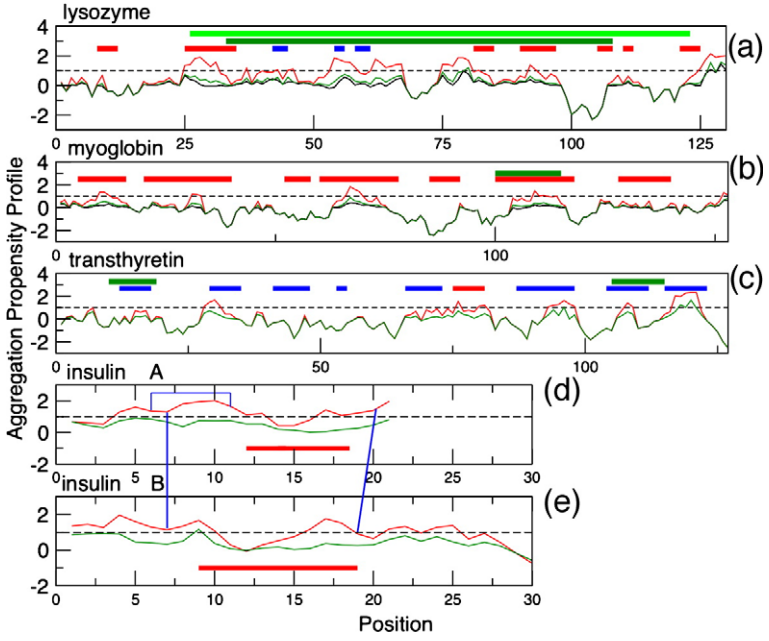

Fig. 2. Predicted aggregation propensity profiles of four structured human proteins. Regions of low structural stability are less protected from aggregation and correspond to the highest peaks in the aggregation propensity profiles $\tilde{Z}_i^{agg}$, which are calculated either by taking into account the structural protection predicted from the sequence (blue lines; see Methods) or from the native state structure (green lines; see Methods). The intrinsic aggregation propensity profiles Z_i^{agg} are reported as red lines. Secondary structure elements are shown as blue bars (β-strands) and red bars (α-helices). (a) Lysozyme: the horizontal bars represent the regions of residues 26–123 and 32–108, which are known to be resistant to proteolysis in the amyloid state.[53] (b) Myoglobin: the horizontal green bar identifies a highly aggregation-prone peptide fragment: residues 100–114.[54] (c) TTR: the regions corresponding to two peptide fragments, residues 10–19[55] and residues 105–115,[56] capable of forming amyloid fibrils are shown as green bars. Insulin, A chain (d) and B chain (e): disulfide bonds are shown by blue lines.

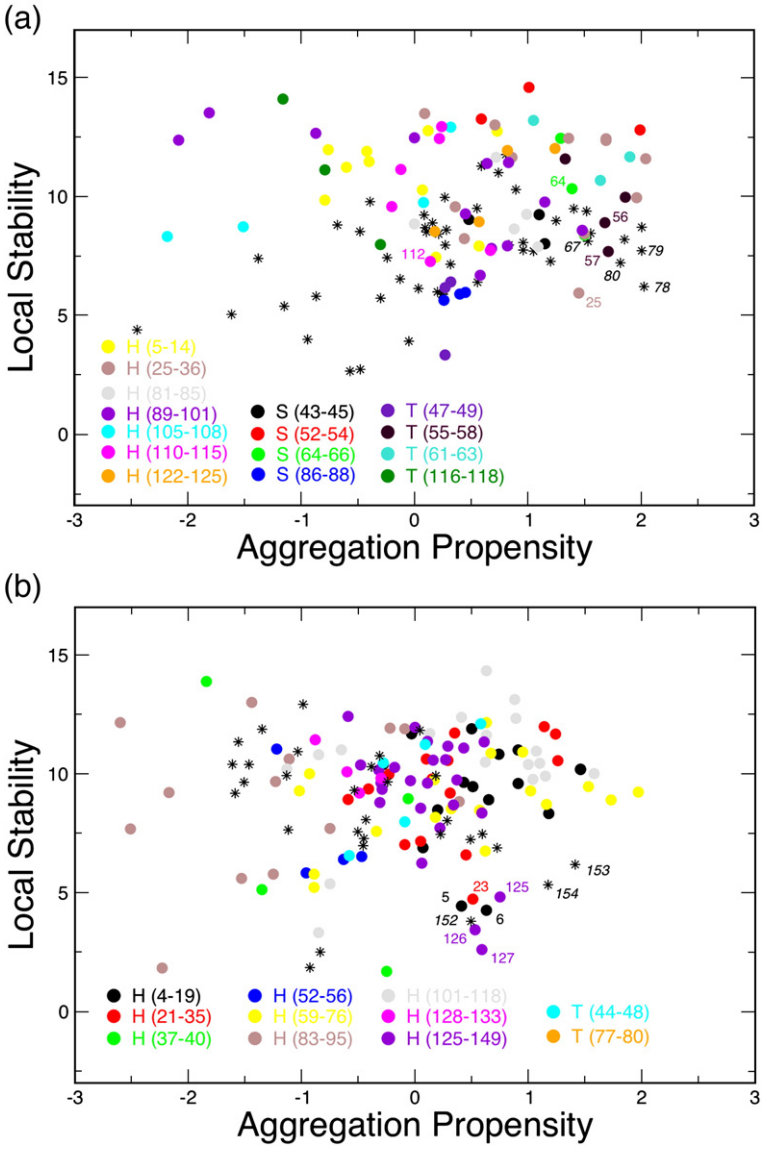

Fig. 3. Relationship between structural protection ($\ln P_i$ score) and intrinsic aggregation propensity (Z_i^{agg} score) at the individual residue level. Residues in different secondary structure elements (H = helix, S = strand, and T = turn) are plotted in different colors; unstructured regions according to ExPASy [www.expasy.org] are marked with stars. (a) Lysozyme. We predict the regions or residues 25–36 (helix), 56–57 (turn), 64–67 (strand), and 78–80 (unstructured) to have simultaneously low structural protection and high aggregation propensity, and therefore to be particularly prone to aggregation under destabilizing conditions; in particular, positions 25 and 78–80 are labeled in bold. We also label in italics the positions associated with amyloidogenic mutations. (b) Myoglobin. We predict amino acids in the regions of residues 4–19 (helix), 21–35 (helix), and 125–149 (helix) to have high aggregation propensity and low structural protection.

Interestingly, residues Ile56, Phe57, Trp64, and Asp67 (labeled in italics in Fig. 3a), which are mutated into Thr, Ile, Arg, and His, respectively, in the four known single-residue familial variants associated with lysozyme amyloidosis,[57] tend to exhibit a combination of high aggregation propensity and low structural stability. It will be interesting to test experimentally the amyloidogenic potential of other residues—in particular Lys25, His78, Lys79, and Ser80—that we identified as having similar characteristics.

Myoglobin

Human myoglobin is a protein non related to disease that has been found *in vitro* to form aggregates sharing all the properties of amyloid fibrils.[58] The aggregation propensity profile $\tilde{Z}_i^{agg}$, calculated by considering the structural protection in the native state, does not exhibit any region above the $\tilde{Z}_i^{agg} = 1$ threshold, in agreement with the fact that myoglobin should be substantially destabilized in order to aggregate.[58] As in the case of lysozyme, very similar $\tilde{Z}_i^{agg}$ profiles are obtained from the calculation of the $\ln P_i$ score from the sequence and from the native structure (blue and green lines, respectively, in Fig. 2b). It is likely that this situation is common for globular proteins, whose native states have evolved to resist aggregation. As with lysozyme, we identified four regions with a high intrinsic aggregation propensity; that is, those above the $Z_i^{agg} = 1$ threshold for the red line in Fig. 2b (residues 9–12, 31–33, 65–70, and 108–114), one of which partially overlaps with a peptide fragment (residues 100–114) found *in vitro* to be highly prone to aggregation.[54] These results suggest that peptide fragments extracted from a full-length protein may exhibit a behavior in their aggregation process different from that of the corresponding regions in the intact polypeptide chain. One of the goals of the current study is to investigate this possibility in terms of the competition between folding and aggregation for specific polypeptide chains.

Similarly to the lysozyme analysis, we compared the aggregation propensity Z_i^{agg} and the structural protection ln P_i at the individual residue level. We predict residues Asp5, Gly6 (helices 4–19), Ala23 (helices 21–35), Gly125, Ala126, and Asp127 (helices 125–149) to have particularly high aggregation propensity and low structural protection. The finding that the regions of the sequence of myoglobin that are most amyloidogenic are not those with a low structural protection in the native state provides another example of how native states of proteins have evolved to be resistant against aggregation.

Transthyretin (TTR)

Amyloid fibrils formed by human TTR are associated with a series of diseases, including familial amyloid polyneuropathy, familial amyloid cardiopathy, central nervous system selective amyloidosis, and senile systemic amyloidosis.[59,60] TTR is a 127-residue protein that adopts a homotetrametic structure[60] in which pairs of TTR monomers, individually resembling a sandwich of two four-stranded β-sheets, come together into a dimer consisting of two eight-stranded intermolecular β-sheets, which can be labeled as DAGHH′G′A′D′ and CBEFF′E′B′C′, respectively.[60] The full TTR tetramer is stabilized by weaker interactions between the two dimers, with contacts occurring through back-to-back hydrophobic interactions of the AB and GH loops.[60] With the use of hydrogen/deuterium exchange detected by NMR spectroscopy, evidence indicating that only two (C, residues 41–49; D, residues 53–55) of the eight native β-strands are exposed to the solvent in the fibrillar structure has been recently presented.[61] We calculated both the intrinsic aggregation propensity profile Z_i^{agg} and the aggregation propensity profile with structural corrections $\tilde{Z}_i^{agg}$ (Fig. 2c). Although the Z_i^{agg} profile exhibits several regions of high intrinsic aggregation propensity (i.e., above the $Z_i^{agg} = 1$ threshold; see Methods), when the structural corrections are considered, a much lower tendency to aggregate is predicted—consistent with the observation that TTR must be destabilized in order to aggregate,[59,60] as we have also noted above for lysozyme and myoglobin.

Insulin

Insulin is a small helical hormone consisting of two polypeptide chains, the A chain (21 residues) and the B chain (30 residues), linked together by two disulfide bridges (A7–B7 and A20–B19), with another disulfide bridge within the A chain (A6–A11).[62] Insulin has been shown to form fibrils under a variety of conditions, including high temperatures and low pH, and in the presence of organic solvents.[63] It has also been shown that both the A chain and the B chain are able to form amyloid fibrils when considered individually.[63] It has also been demonstrated that insulin isomers with alternative disulfide pairings retain their ability to form amyloid fibrils.[62] A systematic study of the aggregation propensities of 6-residue fragments of insulin has revealed that two overlapping peptides from the B chain (residues 11–17 and 12–18) and one peptide from the A chain (residues 13–18) are capable of forming cross-β-structures.[64] We calculated both the intrinsic aggregation propensity profile Z_i^{agg} and the aggregation propensity profile with structural corrections $\tilde{Z}_i^{agg}$. Consistently with experimental results,[62] these profiles indicate that the overall aggregation propensity of the B chain is higher than that of the A chain. As in the case of hIAPP, the presence of disulfide bonds within regions of high intrinsic aggregation propensity appears to prevent them from readily assembling into amyloid deposits.

Prediction of the aggregation-prone regions of prion proteins

Human prion protein (hPrP)

A range of human and animal neurodegenerative diseases—the transmissible spongiform encephalopathies—are associated with the misfolding and aggregation of mammalian prion proteins.[65] The hPrP is involved in sporadic, inherited, or infectious forms of Creutzfeldt–Jakob disease, Gerstmann–Straussler–Sheinker disease, and fatal familial insomnia. The key event in the pathogenesis of these human diseases is the conversion of the normal α-helix-rich and protease-sensitive cellular isoform of the prion protein ($hPrP^C$) into a β-sheet-rich aggregated form of the prion protein ($hPrP^{Sc}$) that possesses distinct physicochemical properties such as protease resistance, insolubility, and potential toxicity.[66] Furthermore, $hPrP^{Sc}$ itself appears to mediate the transmission of transmissible spongiform encephalopathies by promoting the conversion of $hPrP^C$ into its modified and pathogenic aggregated state.

While the mechanism of conversion of $hPrP^C$ into $hPrP^{Sc}$ is not known in detail, specific regions of the $hPrP^C$ sequence appear to be particularly important in modulating interaction with $hPrP^{Sc}$ and in promoting the process of amyloid formation.[28,67] In Fig. 4a, we show the intrinsic aggregation propensity profile Z_i^{agg} for the sequence of hPrP(23–231). We then took into account the effects of the intrinsic propensities of the various residues to be structured, and hence protected, from aggregation resulting in the $\tilde{Z}_i^{agg}$ profile (see Methods). The similarity in the Z_i^{agg} and $\tilde{Z}_i^{agg}$ profiles for residues 23–125 is in agreement with the experimental observation that this region is not structured.[69]

When considering both intrinsic sequence-based propensities and specific structural factors, the region spanning residues 120–126 corresponds to the highest peak in the entire sequence and is the only one to have $\tilde{Z}_i^{agg} > 1$, suggesting that this region is the most aggregation-prone region in the $hPrP^C$ form. This prediction correlates well with experimental data on the *in vitro* aggregation behavior of hPrP fragments. Peptides $hPrP_{106–114}$, $hPrP_{106–126}$, $hPrP_{113–126}$, and $hPrP_{127–147}$ of recombinant hPrP all

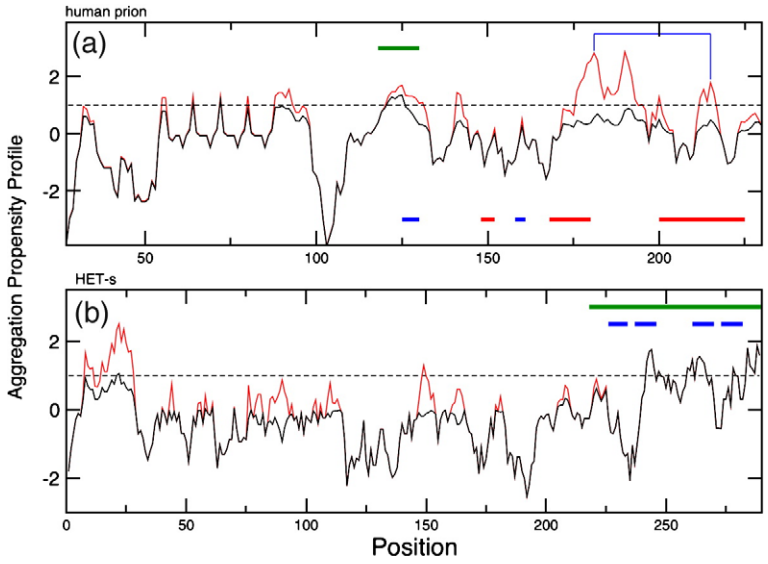

Fig. 4. Aggregation propensity profiles of two prion proteins. Red lines indicate the intrinsic aggregation propensity profiles Z_i^{agg}. Black lines indicate the aggregation propensity profiles $\tilde{Z}_i^{agg}$, calculated by taking into account the structural protection provided by the globular structure of the folded form of the protein. (a) hPrP(23–231). The secondary structure elements present in $hPrP^C$ are indicated as blue bars (β-strands) and red bars (α-helices). The position of the disulfide bond C179–C214 is indicated by a blue line. An experimentally determined aggregation-prone fragment (residues 118–128) is indicated by a green bar, and it is shown to overlap substantially with the major region predicted by our method to have a significant aggregation propensity ($\tilde{Z}_i^{agg}>1$). The region corresponding to the structural core of the amyloid fibril as determined by hydrogen/deuterium exchange (residues 169–213)[68] is indicated by a black bar and corresponds to the region of high intrinsic aggregation propensity ($Z_i^{agg}>1$) formed by residues 175–193. (b) HET-s.[9] The regions corresponding to the four β-strands identified by solid-state NMR are indicated as blue horizontal bars. The green bar indicates the C-terminal fragment whose amyloid structure has been characterized in the latter study.[9]

have high propensities to form amyloid fibrils.[28,67] $hPrP_{106–126}$ has a particularly high intrinsic ability to polymerize into straight and unbranched fibrils and induces apoptosis in primary rat hippocampal cultures,[28] and $hPrP_{113–126}$ is also able to aggregate readily.[67] $hPrP_{106–114}$ and $hPrP_{127–147}$ have a lower tendency to aggregate than $hPrP_{106–126}$, although the former converts into fibrils that are morphologically similar to those formed by $hPrP_{106–126}$, while the latter forms twisted fibrillary structures.[67] A recent report has identified two other peptide fragments, $hPrP_{119–126}$ and $hPrP_{121–127}$, that can readily form amyloid-like fibrils and can be cytotoxic to astrocytes.[70] These fragments include, at least in part, the region 118–128 of the sequence (Fig. 4a). Further support for our predictions comes from studies of another mammalian prion protein. A deletion mutant of recombinant mouse PrP (moPrPC Δ114–121) does not readily aggregate and indeed is able to inhibit the conversion of endogenous wild-type PrPC into PrPSc in scrapie-infected mouse neuroblastoma cells.[71] Since the sequences of moPrP and hPrP are very similar (90% sequence identity, and 100% sequence identity in the 114–121 region), this result emphasizes the importance of the deleted segment for the aggregation of mammalian prion proteins.

We also calculated a significant propensity for aggregation in the vicinity of the copper-binding region comprising the four tandem repeats of the octapeptide sequence PHGGGWGQ, in agreement with the observation that this region may play an important role in the oligomerization process of this protein.[72]

A central result of the present work concerns the aggregation propensity of the region 175–193, which includes α-helix II in the $hPrP^C$ form. The aggregation propensity profile Z_i^{agg} predicted by considering only the intrinsic physicochemical factors (Fig. 4a) identifies this region as the most amyloidogenic one. However, using the CamP method (see Methods), we also predicted, in agreement with experimental data,[69] that this region is highly structured in the $hPrP^C$ form. Therefore, when the aggregation propensity profile $\tilde{Z}_i^{agg}$ is considered, the region of residues 175–193 becomes less aggregation-prone in the $hPrP^C$ form than the region of residues 118–128. In addition, the presence of the disulfide bond C179–C214 appears to play an important role in stabilizing the region 175–193 in the $hPrP^C$ form by inhibiting the formation of intermolecular interactions from this state.[73] However, recent hydrogen/deuterium exchange experiments[68] have indicated that the region corresponding to the structural core of the amyloid fibril corresponds to residues 169–213[68] (black bar in Fig. 4a)—a result that is in close agreement with our predictions using the intrinsic aggregation propensity profile Z_i^{agg}. Therefore, the comparison of the Z_i^{agg} and $\tilde{Z}_i^{agg}$ profiles suggests that the region of residues 175–193 is involved in the conformational conversion from the $hPrP^C$ form into the $hPrP^{Sc}$ form.

The comparison of the aggregation propensity (Z_i^{agg} score) and the structural stability ($\ln P_i$ score) at the individual residue level is shown in Fig. 5. We observe that the region of residues 121–125 has simultaneously the highest aggregation propensity and the lowest structural protection, thus extending the discussion presented above about the high potential for aggregation of the region of residues 120–126.

HET-s

HET-s of the yeast *Podospora anserina* is a prion protein that is involved in heterokaryon incompatibility and is not associated with disease.[74] HET-s

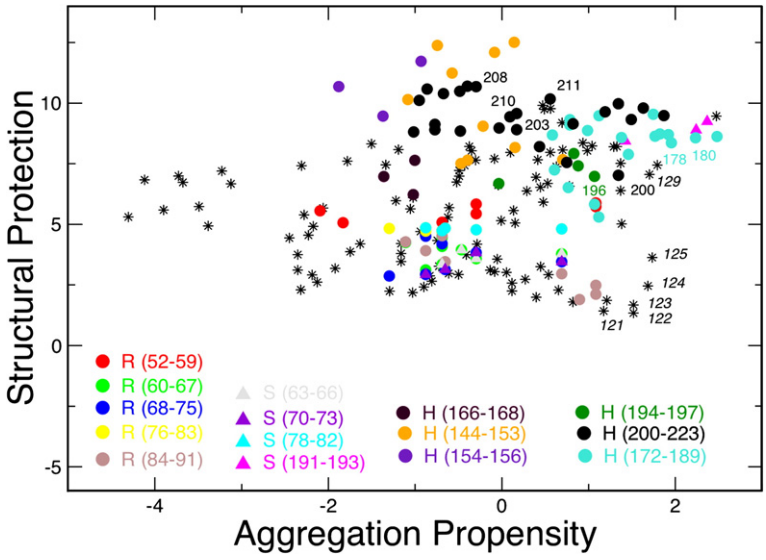

Fig. 5. Relationship between folding ($\ln P_i$ score) and intrinsic aggregation (Z_i^{agg} score) propensities at the individual residue level for the hPrP. Residues in different secondary structure elements (H = helix, S = strand, and T = turn) are plotted in different colors. Unstructured regions according to ExPASy [www.expasy.org] are marked with stars. Residues 120–126, which have particularly high intrinsic aggregation propensity and low structural protection, are labeled in bold.

has been shown to form amyloid fibrils,[75] whose structures have been characterized through solid-state NMR, in conjunction with site-directed fluorescence labeling and hydrogen exchange protocol.[9] In the resulting structural model of the fibrils from the C-terminal fragment of HET-s (residues 218–289), each molecule contributes four β-strands, with strands 1 and 3 (residues 226–234 and 262–270) forming a parallel β-sheet and with strands 2 and 4 (residues 237–245 and 273–282) forming another parallel β-sheet located about 10 Å away.[9] These β-strands are connected by two short loops between β1 and β2, and between β3 and β4, respectively, and by an unstructured 15-residue segment between β2 and β3.

Calculation of the intrinsic aggregation propensity profile Z_i^{agg} reveals a high aggregation propensity in the regions of residues 5–22 and 245–289 (Fig. 4b). The monomeric form of HET-s appears to be structured in the region of residues 1–227 and rather unstructured in the region of residues 228–289.[9] Consistent with these results, we determined a much lower propensity for aggregation in the N-terminal region through the $\tilde{Z}_i^{agg}$ profile (Fig. 4b), which results in part from the very high structural protection predicted for this region by the CamP method (see Methods). The region encompassing residues 228–289 is therefore predicted to be the principal aggregation-prone region, being the only one with $\tilde{Z}_i^{agg}$. This fragment, in contrast with the fragment 1–227, retains the ability to form fibrils *in vitro*, catalyzes efficiently the aggregation of full-length HET-s, and is able to induce prion propagation *in vivo*.[76] In addition, limited proteolysis experiments indicate that the region of residues 218–289 is in the fibril core.[76] Three of the four β-strands identified experimentally (residues 226–234, 237–244, 262–271, and 273–282) as those forming the core of the cross-β-structure correspond to the three major peaks (residues 242–245, 260–267, and 278–289) in the aggregation propensity profile $\tilde{Z}_i^{agg}$ of HET-s (Fig. 4b). We therefore suggest that while β1 appears to form the β-core in the amyloid fibril, it is not likely to be directly involved in the process of aggregation. A structural disturbance has actually been found in the fibrils at the level of β1,[9] suggesting that this region may have a lower propensity, relative to the other strands, to drive amyloid fibril formation.

Conclusions

We have described in this article a method for predicting the regions of the sequences of proteins that are most important in promoting their aggregation from their structured, partially structured, or unstructured state. Our analysis has revealed that the regions that promote aggregation, even from globular states, can be identified based on the knowledge of the amino acid sequence. We have also found that the regions of high intrinsic aggregation propensity tend to be highly protected from aggregation in the native states of globular proteins, indicating the existence of evolutionary strategies to resist aggregation in these states. In order to promote the aggregation process, these regions should be destabilized so that they become available to form intermolecular interactions.

The methodology that we have presented is general and is based on the idea that the sequence of a protein determines its behavior in the case of both folding[77] and misfolding.[17,78] The possibility of predicting the aggregation-promoting regions for natively unfolded polypeptide chains, for globular proteins, and for systems that contain both folded and unfolded domains—as provided by methods such as the one that we have presented here—should be of significant value in developing rational approaches to the avoidance of aggregation in biotechnology and to the treatment of protein deposition diseases, as it identifies the major factors determining aggregation and the regions in which these factors are prevalent.

Methods

Intrinsic propensity profiles for aggregation of polypeptide sequences

In the approach described in this article, the intrinsic aggregation propensities of individual amino acids are defined as:[1]

$$p_i^{\mathrm{agg}} = \alpha_{\mathrm{h}} p_i^{\mathrm{h}} + \alpha_{\mathrm{s}} p_i^{\mathrm{s}} + \alpha_{\mathrm{hyd}} p_i^{\mathrm{hyd}} + \alpha_{\mathrm{c}} p_i^{\mathrm{c}} \quad (1)$$

where p_i^{h} and p_i^{s} are the propensities for α-helix and β-sheet formation, respectively, of the amino acid at position i; and p_i^{hyd} and p_i^{c} are the hydrophobicity and charge, respectively. These propensities are then combined in a linear manner with coefficients α determined as described below. The p_i^{agg} values are combined to provide a profile P_i^{agg}, which describes the intrinsic propensity for aggregation as a function of the complete amino acid sequence.[1] At each position i along the sequence, we define the profile P_i^{agg} as an average over a window of seven residues:

$$P_i^{\mathrm{agg}} = \frac{1}{7}\sum_{j=-3}^{3} p_{i+j}^{\mathrm{agg}} + \alpha_{\mathrm{pat}} I_i^{\mathrm{pat}} + \alpha_{\mathrm{gk}} I_i^{\mathrm{gk}} \quad (2)$$

where I_i^{pat} is the term that takes into account the presence of specific patterns of alternating hydrophobic and hydrophilic residues, strongly promoting aggregation,[1,79] and I_i^{gk} is the term that takes into account the gatekeeping effect of individual charges c_i,[80,81] modulating the effects of the hydrophobic–hydrophilic patterns as follows:

$$I_i^{\mathrm{gk}} = \sum_{j=-10}^{10} c_{i+j} \quad (3)$$

In this work, in the sums involved in sliding windows, we considered shorter windows at the N- and C-termini. The parameters α were fitted in accordance with a procedure similar to the one described by Dubay *et al.*[17] In order to compare the intrinsic propensity profiles, we normalize P_i^{agg} by considering the average μ^{agg} and the standard deviation σ^{agg} of P_i^{agg} at each position i for random sequences:

$$Z_i^{\mathrm{agg}} = \frac{P_i^{\mathrm{agg}} - \mu^{\mathrm{agg}}}{\sigma^{\mathrm{agg}}} \quad (4)$$

where the average μ^{agg} and the standard deviation σ^{agg} are calculated over random sequences as:

$$\mu^{\mathrm{agg}} = \frac{1}{(N-6)N_S}\sum_{k=1}^{N_S}\sum_{i=4}^{N-4} Z_i^{\mathrm{agg}}(S_k)$$
$$\sigma^{\mathrm{agg}} = \sqrt{\frac{1}{(N-6)N_S}\sum_{k=1}^{N_S}\sum_{i=4}^{N-4}\left(Z_i^{\mathrm{agg}}(S_k) - \mu^{\mathrm{agg}}\right)^2} \quad (5)$$

In these formulas, we considered N_S as the random sequences of length N, and we verified that μ^{agg} and σ^{agg} are nearly constant for values of N ranging from 50 to 1000. Random sequences were generated by using the amino acid frequencies of the SWISS-PROT database.[82] The plot of Z_i^{agg} *versus* the residue number (from residue 1 to residue N) represents the intrinsic aggregation propensity profile, as reported.[1]

Prediction of folding propensities from the sequence

We used the CamP method,[25] by which the flexibility and the solvent accessibility of proteins are predicted with high accuracy. The CamP method provides a position-dependent score, denoted as $\ln P_i$,[25] which predicts the local stability of the structure in the region of residue i. This method enables predictions from the knowledge of amino acid sequences of the buried regions with >80% accuracy and of the protection factors for hydrogen exchange with an average accuracy of 60%.[25]

Prediction of aggregation propensity profiles for partially structured polypeptide chains

A region of a polypeptide sequence should meet two conditions in order to promote aggregation: (1) it should have a high intrinsic aggregation propensity ($Z_i^{\mathrm{agg}} > Z_{\mathrm{threshold}}^{\mathrm{agg}}$), and (2) it should be sufficiently unstructured or unstable to have a significant propensity to form intermolecular interactions. In order to describe the latter, we use the CamP method, which provides a prediction of the protection factors from hydrogen exchange $\ln P_i$ by using the knowledge of the amino acid sequence of a protein.[25] Values $Z_i^{\mathrm{agg}} > 0$ are modified by modulating them with $\ln P_i$:

$$\tilde{Z}_i^{\mathrm{agg}} = Z_i^{\mathrm{agg}}(1 - \varepsilon \ln P_i) \quad (6)$$

where $\varepsilon = 1/15$ was chosen to have a modulation factor in the range between 0 and 1. The plot of $\tilde{Z}_i^{\mathrm{agg}}$ *versus* residue number (from residue 1 to residue N) represents the aggregation propensity profile calculated to account for the structural protection. We note that, in the calculation of $\tilde{Z}_i^{\mathrm{agg}}$, we considered $Z_{\mathrm{threshold}}^{\mathrm{agg}} = 0$, while in the plots of the aggregation propensity profiles, we considered $Z_{\mathrm{threshold}}^{\mathrm{agg}} = 1$.

The protection factors can be predicted from the structure of a protein when the latter is known.[83] In this case, we apply Eq. (6) by using the values $\ln P_i$ calculated according to the equation:

$$\ln P_i = \beta_{\mathrm{c}} N_{\mathrm{c}} + \beta_{\mathrm{h}} N_{\mathrm{h}} \quad (7)$$

where N_{c} and N_{h} are, respectively, the number of interatomic contacts made by the amide nitrogen and the number of hydrogen bonds made by the amide hydrogen of residue i, and $\beta_{\mathrm{c}} = 0.35$ and $\beta_{\mathrm{h}} = 2$ are two parameters chosen by a fitting procedure to optimize the agreement between predicted and experimental protection factors for a database of eight proteins.[83] In the calculations presented in this work, we have used the following Protein Data Bank files: 1sjb (lysozyme), 2mm1 (myoglobin), 1tz8 (TTR), and 1wht (insulin).

We finally point out that when the protection factors are known experimentally, they can be used directly in Eq. (6). We also note that the structural protection provided by the presence of disulfide bonds is not currently taken directly into account through Eq. (6).

Absolute propensities for aggregation of structured polypeptide sequences

Only residues with high $\tilde{Z}_i^{\mathrm{agg}}$ values ($Z_{\mathrm{threshold}}^{\mathrm{agg}} = 0$) are considered as contributing to the aggregation propensity $\tilde{Z}_i^{\mathrm{agg}}$, resulting in the formula:

$$\tilde{Z}^{\mathrm{agg}} = \frac{\sum_{i=1}^{N} \tilde{Z}_i^{\mathrm{agg}} \vartheta\left(\tilde{Z}_i^{\mathrm{agg}}\right)}{\sum_{i=1}^{N} \vartheta\left(\tilde{Z}_i^{\mathrm{agg}}\right)} \quad (8)$$

where the function $\vartheta(\tilde{Z}_i^{agg})$ is 1 for $\vartheta(\tilde{Z}_i^{agg}) \geq 0$, and 0 for $\vartheta(\tilde{Z}_i^{agg}) < 0$. We use a similar expression for computing the absolute aggregation propensity without structural corrections:

$$Z^{agg} = \frac{\sum_{i=1}^{N} Z_i^{agg} \vartheta(Z_i^{agg})}{\sum_{i=1}^{N} \vartheta(Z_i^{agg})} \quad (9)$$

Software availability

The Zyggregator method for predicting the aggregation propensities of peptides and proteins and the CamP method for predicting the flexibility and solvent accessibility of protein structures are available at the Vendruscolo group web site†.

Acknowledgements

This work was supported by the Leverhulme Trust (G.G.T., C.M.D., and M.V.), the European Union (C. M.D. and M.V.), the Wellcome Trust (C.M.D.), and the Royal Society (M.V.).

References

1. Pawar, A. P., DuBay, K. F., Zurdo, J., Chiti, F., Vendruscolo, M. & Dobson, C. M. (2005). Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases. *J. Mol. Biol.* **350**, 379–392.
2. de Groot, N. S., 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.
3. 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.
4. Chiti, F., Taddei, N., Baroni, F., Capanni, C., Stefani, M., Ramponi, G. & Dobson, C. M. (2002). Kinetic partitioning of protein folding and aggregation. *Nat. Struct. Biol.* **9**, 137–143.
5. Heise, H., Hoyer, W., Becker, S., Andronesi, O. C., Riedel, D. & Baldus, M. (2005). Molecular-level secondary structure, polymorphism, and dynamics of full-length alpha-synuclein fibrils studied by solid-state NMR. *Proc. Natl Acad. Sci. USA*, **102**, 15871–15876.
6. Kajava, A. V., Aebi, U. & Steven, A. C. (2005). The parallel superpleated beta-structure as a model for amyloid fibrils of human amylin. *J. Mol. Biol.* **348**, 247–252.
7. Krishnan, R. & Lindquist, S. L. (2005). Structural insights into a yeast prion illuminate nucleation and strain diversity. *Nature*, **435**, 765–772.
8. 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.
9. Ritter, C., Maddelein, M. L., Siemer, A. B., Luhrs, T., Ernst, M., Meier, B. H. *et al.* (2005). Correlation of structural elements and infectivity of the HET-s prion. *Nature*, **435**, 844–848.
10. Sawaya, M. R., Sambashivan, S., Nelson, R., Ivanova, M. I., Sievers, S. A., Apostol, M. I. *et al.* (2007). Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature*, **447**, 453–457.
11. Fowler, S. B., Poon, S., Muff, R., Chiti, F., Dobson, C. M. & Zurdo, J. (2005). Rational design of aggregation-resistant bioactive peptides: reengineering human calcitonin. *Proc. Natl Acad. Sci. USA*, **102**, 10105–10110.
12. Ventura, S. & Villaverde, A. (2006). Protein quality in bacterial inclusion bodies. *Trends Biotechnol.* **24**, 179–185.
13. Chalifour, R. J., McLaughlin, R. W., Lavoie, L., Morissette, C., Tremblay, N., Boule, M. *et al.* (2003). Stereoselective interactions of peptide inhibitors with the beta-amyloid peptide. *J. Biol. Chem.* **278**, 34874–34881.
14. Soto, C., Sigurdsson, E. M., Morelli, L., Kumar, R. A., Castano, E. M. & Frangione, B. (1998). Beta-sheet breaker peptides inhibit fibrillogenesis in a rat brain model of amyloidosis: implications for Alzheimer's therapy. *Nat. Med.* **4**, 822–826.
15. Tatarek-Nossol, M., Yan, L. M., Schmauder, A., Tenidis, K., Westermark, G. & Kapurniotu, A. (2005). Inhibition of NAPP amyloid-fibril formation and apoptotic cell death by a designed hIAPP amyloid-core-containing hexapeptide. *Chem. Biol.* **12**, 797–809.
16. 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.
17. 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.
18. Rousseau, F., Schymkowitz, J. & Serrano, L. (2006). Protein aggregation and amyloidosis: confusion of the kinds? *Curr. Opin. Struct. Biol.* **16**, 118–126.
19. 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.
20. Thompson, M. J., Sievers, S. A., Karanicolas, J., Ivanova, M. I., Baker, D. & Eisenberg, D. (2006). The 3D profile method for identifying fibril-forming segments of proteins. *Proc. Natl Acad. Sci. USA*, **103**, 4074–4078.
21. Trovato, A., Chiti, F., Maritan, A. & Seno, F. (2006). Insight into the structure of amyloid fibrils from the analysis of globular proteins. *PLoS Comp. Biol.* **2**, 1608–1618.
22. Galzitskaya, O. V., Garbuzynskiy, S. O. & Lobanov, M. Y. (2006). Prediction of amyloidogenic and disordered regions in protein chains. *PLoS Comp. Biol.* **2**, 1639–1648.
23. Conchillo-Sole, O., de Groot, N. S., Aviles, F. X., Vendrell, J., Daura, X. & Ventura, S. (2007). AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BMC Bioinf.* **8**, 65.
24. Zibaee, S., Makin, O. S., Goedert, M. & Serpell, L. C. (2007). A simple algorithm locates beta-strands in the amyloid fibril core of alpha-synuclein, A beta, and tau

† http://www-vendruscolo.ch.cam.ac.uk/software.html

using the amino acid sequence alone. *Protein Sci.* **16**, 906–918.
25. Tartaglia, G. G., Cavalli, A. & Vendruscolo, M. (2007). Prediction of local structural stabilities of proteins from their amino acid sequences. *Structure*, **15**, 139–143.
26. 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.
27. Meinhardt, J., Tartaglia, G. G., Pawar, A., Christopeit, T., Hortschansky, P., Schroeckh, V. *et al.* (2007). Similarities in the thermodynamics and kinetics of aggregation of disease-related A beta(1–40) peptides. *Protein Sci.* **16**, 1214–1222.
28. Forloni, G., Angeretti, N., Chiesa, R., Monzani, E., Salmona, M., Bugiani, O. & Tagliavini, F. (1993). Neurotoxicity of a prion protein fragment. *Nature*, **362**, 543–546.
29. Reches, M., Porat, Y. & Gazit, E. (2002). Amyloid fibril formation by pentapeptide and tetrapeptide fragments of human calcitonin. *J. Biol. Chem.* **277**, 35475–35480.
30. Luhrs, T., Ritter, C., Adrian, M., Riek-Loher, D., Bohrmann, B., Doeli, H. *et al.* (2005). 3D structure of Alzheimer's amyloid-beta(1–42) fibrils. *Proc. Natl Acad. Sci. USA*, **102**, 17342–17347.
31. 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.
32. Moriarty, D. F., Vagts, S. & Raleigh, D. P. (1998). A role for the C-terminus of calcitonin in aggregation and gel formation: a comparative study of C-terminal fragments of human and salmon calcitonin. *Biochem. Biophys. Res. Commun.* **245**, 344–348.
33. Pedersen, J. S., Dikov, D. & Otzen, D. E. (2006). N- and C-terminal hydrophobic patches are involved in fibrillation of glucagon. *Biochemistry*, **45**, 14503–14512.
34. Ferguson, N., Becker, J., Tidow, H., Tremmel, S., Sharpe, T. D., Krause, G. *et al.* (2006). General structural motifs of amyloid protofilaments. *Proc. Natl Acad. Sci. USA*, **103**, 16248–16253.
35. Luca, S., Yau, W. M., Leapman, R. & Tycko, R. (2007). Peptide conformation and supramolecular organization in amylin fibrils: constraints from solid-state NMR. *Biochemistry*, **46**, 13505–13522.
36. 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.
37. Del Mar, C., Greenbaum, E. A., Mayne, L., Englander, S. W. & Woods, V. L. (2005). Structure and properties of alpha-synuclein and other amyloids determined at the amino acid level. *Proc. Natl Acad. Sci. USA*, **102**, 15477–15482.
38. Miake, H., Mizusawa, H., Iwatsubo, T. & Hasegawa, M. (2002). Biochemical characterization of the core structure of alpha-synuclein filaments. *J. Biol. Chem.* **277**, 19213–19219.
39. Selkoe, D. J. (2002). Alzheimer's disease is a synaptic failure. *Science*, **298**, 789–791.
40. Petkova, A. T., Yau, W. M. & Tycko, R. (2006). Experimental constraints on quaternary structure in Alzheimer's beta-amyloid fibrils. *Biochemistry*, **45**, 498–512.
41. Westermark, P. (1994). Amyloid and polypeptide hormones—what is their interrelationship? *Amyloid*, **1**, 47–60.
42. Cudd, A., Arvinte, T., Das, R. E. G., Chinni, C. & Macintyre, A. (1995). Enhanced potency of human calcitonin when fibrillation is avoided. *J. Pharm. Sci.* **84**, 717–719.
43. Kamihira, M., Naito, A., Tuzi, S., Nosaka, A. Y. & Saito, H. (2000). Conformational transitions and fibrillation mechanism of human calcitonin as studied by high-resolution solid-state C-13 NMR. *Protein Sci.* **9**, 867–877.
44. Cryer, P. E., Davis, S. N. & Shamoon, H. (2003). Hypoglycemia in diabetes. *Diabetes Care*, **26**, 1902–1912.
45. Ferguson, N., Berriman, J., Petrovich, M., Sharpe, T. D., Finch, J. T. & Fersht, A. R. (2003). Rapid amyloid fiber formation from the fast-folding WW domain FBP28. *Proc. Natl Acad. Sci. USA*, **100**, 9814–9819.
46. Hull, R. L., Westermark, G. T., Westermark, P. & Kahn, S. E. (2004). Islet amyloid: a critical entity in the pathogenesis of type 2 diabetes. *J. Clin. Endocrinol. Metab.* **89**, 3629–3643.
47. Scrocchi, L. A., Chen, Y., Waschuk, S., Wang, F., Cheung, S., Darabie, A. A. *et al.* (2002). Design of peptide-based inhibitors of human islet amyloid polypeptide fibrillogenesis. *J. Mol. Biol.* **318**, 697–706.
48. Yan, L. M., Tatarek-Nossol, M., Velkova, A., Kazantzis, A. & Kapurniotu, A. (2006). Design of a mimic of nonamyloidogenic and bioactive human islet amyloid polypeptide (IAPP) as nanomolar affinity inhibitor of IAPP cytotoxic fibrillogenesis. *Proc. Natl Acad. Sci. USA*, **103**, 2046–2051.
49. Porat, Y., Mazor, Y., Efrat, S. & Gazit, E. (2004). Inhibition of islet amyloid polypeptide fibril formation: a potential role for heteroaromatic interactions. *Biochemstry*, **43**, 14454–14462.
50. Pepys, M. B., Hawkins, P. N., Booth, D. R., Vigushin, D. M., Tennent, G. A., Soutar, A. K. *et al.* (1993). Human lysozyme gene-mutations cause hereditary systemic amyloidosis. *Nature*, **362**, 553–557.
51. Valleix, S., Drunat, S., Philit, J. B., Adoue, D., Piette, J. C., Droz, D. *et al.* (2002). Hereditary renal amyloidosis caused by a new variant lysozyme W64R in a French family. *Kidney Int.* **61**, 907–912.
52. Yazaki, M., Farrell, S. A. & Benson, M. D. (2003). A novel lysozyme mutation Phe57Ile associated with hereditary renal amyloidosis. *Kidney Int.* **63**, 1652–1657.
53. Frare, E., Mossuto, M. F., de Laureto, P. P., Dumoulin, M., Dobson, C. M. & Fontana, A. (2006). Identification of the core structure of lysozyme amyloid fibrils by proteolysis. *J. Mol. Biol.* **361**, 551–561.
54. Fandrich, M., Forge, V., Buder, K., Kittler, M., Dobson, C. M. & Diekmann, S. (2003). Myoglobin forms amyloid fibrils by association of unfolded polypeptide segments. *Proc. Natl Acad. Sci. USA*, **100**, 15463–15468.
55. Jarvis, J. A., Kirkpatrick, A. & Craik, D. J. (1994). H-1 NMR analysis of fibril-forming peptide fragments of transthyretin. *Int. J. Pept. Protein Res.* **44**, 388–398.
56. Jaroniec, C. P., MacPhee, C. E., Bajaj, V. S., McMahon, M. T., Dobson, C. M. & Griffin, R. G. (2004). High-resolution molecular structure of a peptide in an amyloid fibril determined by magic angle spinning NMR spectroscopy. *Proc. Natl Acad. Sci. USA*, **101**, 711–716.
57. Dumoulin, M., Kumita, J. R. & Dobson, C. M. (2006). Normal and aberrant biological self-assembly: insights from studies of human lysozyme and its amyloidogenic variants. *Acc. Chem. Res.* **39**, 603–610.
58. Fandrich, M., Fletcher, M. A. & Dobson, C. M. (2001). Amyloid fibrils from muscle myoglobin—even an ordinary globular protein can assume a rogue guise if conditions are right. *Nature*, **410**, 165–166.

59. Hammarstrom, P., Wiseman, R. L., Powers, E. T. & Kelly, J. W. (2003). Prevention of transthyretin amyloid disease by changing protein misfolding energetics. *Science*, **299**, 713–716.
60. Laidman, J., Forse, G. J. & Yeates, T. O. (2006). Conformational change and assembly through edge beta strands in transthyretin and other amyloid proteins. *Acc. Chem. Res.* **39**, 576–583.
61. Olofsson, A., Ippel, J. H., Wijmenga, S. S., Lundgren, E. & Ohman, A. (2003). Probing solvent accessibility of transthyretin amyloid by solution NMR spectroscopy. *J. Biol. Chem.* **279**, 5699–5707.
62. Huang, K., Maiti, N. C., Phillips, N. B., Carey, P. R. & Weiss, M. A. (2006). Structure-specific effects of protein topology on cross-beta assembly: studies of insulin fibrillation. *Biochemistry*, **45**, 10278–10293.
63. Hong, D. P., Ahmad, A. & Fink, A. L. (2006). Fibrillation of human insulin A and B chains. *Biochemistry*, **45**, 9342–9353.
64. Ivanova, M. I., Thompson, M. J. & Eisenberg, D. (2006). A systematic screen of beta(2)-microglobulin and insulin for amyloid-like segments. *Proc. Natl Acad. Sci. USA*, **103**, 4079–4082.
65. Prusiner, S. B., McKinley, M. P., Bowman, K. A., Bolton, D. C., Bendheim, P. E., Groth, D. F. & Glenner, G. G. (1983). Scrapie prions aggregate to form amyloid-like birefringent rods. *Cell*, **35**, 349–358.
66. Prusiner, S. B. (1991). Molecular biology of prion diseases. *Science*, **252**, 1515–1522.
67. Tagliavini, F., Prelli, F., Verga, L., Giaccone, G., Sarma, R., Gorevic, P. *et al.* (1993). Synthetic peptides homologous to prion protein residues 106–147 form amyloid-like fibrils *in-vitro*. *Proc. Natl Acad. Sci. USA*, **90**, 9678–9682.
68. Lu, X. J., Wintrode, P. L. & Surewicz, W. K. (2007). Beta-sheet core of human prion protein amyloid fibrils as determined by hydrogen/deuterium exchange. *Proc. Natl Acad. Sci. USA*, **104**, 1510–1515.
69. Zahn, R., Liu, A. Z., Luhrs, T., Riek, R., von Schroetter, C., Garcia, F. L. *et al.* (2000). NMR solution structure of the human prion protein. *Proc. Natl Acad. Sci. USA*, **97**, 145–150.
70. Satheeshkumar, K. S., Murali, J. & Jayakumar, R. (2004). Assemblages of prion fragments: novel model systems for understanding amyloid toxicity. *J. Struct. Biol.* **148**, 176–193.
71. Holscher, C., Delius, H. & Burkle, A. (1998). Overexpression of nonconvertible PrPc Delta 114–121 in scrapie-infected mouse neuroblastoma cells leads to trans-dominant inhibition of wild-type PrPSc accumulation. *J. Virol.* **72**, 1153–1159.
72. Wells, M. A., Jackson, G. S., Jones, S., Hosszu, L. L. P., Craven, C. J., Clarke, A. R. *et al.* (2006). A reassessment of copper(II) binding in the full-length prion protein. *Biochem. J.* **399**, 435–444.
73. Hosszu, L. L. P., Baxter, N. J., Jackson, G. S., Power, A., Clarke, A. R., Waltho, J. P. *et al.* (1999). Structural mobility of the human prion protein probed by backbone hydrogen exchange. *Nat. Struct. Biol.* **6**, 740–743.
74. Coustou, V., Deleu, C., Saupe, S. & Begueret, J. (1997). The protein product of the HET-s heterokaryon incompatibility gene of the fungus *Podospora anserina* behaves as a prion analog. *Proc. Natl Acad. Sci. USA*, **94**, 9773–9778.
75. Dos Reis, S., Coulary-Salin, B., Forge, V., Lascu, I., Begueret, J. & Saupe, S. J. (2002). The HET-s prion protein of the filamentous fungus *Podospora anserina* aggregates *in vitro* into amyloid-like fibrils. *J. Biol. Chem.* **277**, 5703–5706.
76. Balguerie, A., Dos Reis, S., Ritter, C., Chaignepain, S., Coulary-Salin, B., Forge, V. *et al.* (2003). Domain organization and structure–function relationship of the HET-s prion protein of *Podospora anserina*. *EMBO J.* **22**, 2071–2081.
77. Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science*, **181**, 223–230.
78. Dobson, C. M. (1999). Protein misfolding, evolution and disease. *Trends Biochem. Sci.* **24**, 329–332.
79. 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. *J. Mol. Biol.* **96**, 11211–11216.
80. Otzen, D. E., Kristensen, O. & Oliveberg, M. (2000). Salt-induced detour through compact regions of the protein folding landscape. *Proc. Natl Acad. Sci. USA*, **97**, 9907–9912.
81. Rousseau, F., Serrano, L. & Schymkowitz, J. W. H. (2004). How evolutionary pressure against protein aggregation shaped chaperone specificity. *J. Mol. Biol.* **355**, 1037–1047.
82. Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E. *et al.* (2003). The SWISS-PROT protein knowledge base and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **31**, 365–370.
83. Best, R. B. & Vendruscolo, M. (2006). Structural interpretation of hydrogen exchange protection factors in proteins: characterization of the native state fluctuations of C12. *Structure*, **14**, 97–106.