Sequence and Structural Determinants of Amyloid Fibril Formation

FRANCESCO BEMPORAD, GIULIA CALLONI, SILVIA CAMPIONI, GEORGIA PLAKOUTSI, NICCOLÒ TADDEI, AND FABRIZIO CHITI*

Dipartimento di Scienze Biochimiche, Università di Firenze, Viale Morgagni 50, 50134 Firenze, Italy

Received January 4, 2006

ABSTRACT

Amyloid fibril formation is a process that represents an essential feature of the chemistry of proteins and plays a central role in human pathology and the biology of living organisms. In this Account, we shall describe some of the recent results on the sequence and structural determinants of protein aggregation. We shall describe the factors that govern aggregation of unfolded peptides and proteins. We shall then try to summarize the factors that pertain to the aggregation of partially structured states and will show that even fully folded states of proteins have an ability to aggregate into at least early oligomers with no need to undergo substantial conformational changes.

Introduction

Proteins and peptides have a general tendency to convert from their soluble states into well-organized aggregates characterized by a fibrillar morphology and an extended cross-β structure.[1,2] The generic potential of polypeptide chains to generate these fibrillar aggregates, generally referred to as amyloid fibrils, is relevant for a number of reasons. First, it represents an essential feature of the behavior of polypeptide chains that needs to be fully understood for a thorough characterization of the dynamics and conformational changes of proteins. Second, amyloid aggregates, especially in the form of early forming oligomers, have an inherent toxicity to cells.[3] The elucidation of the early steps of protein aggregation facilitates an understanding of those housekeeping functions that exist in any living organism to prevent the formation of such detrimental species. Third, the formation of amyloid fibrils or intracellular inclusions with amyloid-like characteristics is associated with over 40 pathological conditions in humans,[3] all having distinct and well-described clinical profiles. Finally, such fibrillar species can serve a number of biological functions in living organisms, provided that they form under controlled conditions.[3,4] Perhaps the most fascinating of these functions lies in the ability of amyloid-like fibrils to serve as transmissible genetic traits distinct from DNA genes.[5]

Investigating the mechanism of amyloid fibril formation therefore involves shedding light on a process that represents an essential feature of the chemistry of proteins, plays a central role in human pathology, and constitutes an important component of the biology of living organisms. The achievement of this goal implies identifying all of the conformational states and oligomeric structures adopted by the polypeptide chain in the overall process and the thermodynamics and kinetics of all of the conformational changes between them. It also implies identifying the residues or regions of the sequence that promote each of these steps. In this Account, we shall try to describe some of the recent advances that have contributed to increase our knowledge of the sequence and structural determinants of protein aggregation.

Aggregation of Unfolded Polypeptide Chains Is Governed by Simple Physicochemical Factors. First Observations

Many proteins, including several involved in protein deposition diseases, are normally largely unstructured. Proteins characterized by a globular structure can also experience unfolded conformations during their lifespan in the cell, for example, during and immediately after translation, trafficking, or under stress conditions.[6] An investigation of the aggregation process of unfolded human muscle acylphosphatase (AcP) has provided one of the first suggestions that amyloid fibril formation of unfolded systems could be governed by relatively simple and general rules.[7,8] After denaturation in the presence of moderate concentrations of trifluoroethanol (TFE), AcP was shown to aggregate into structures with a morphology and size typical of amyloid protofilaments and fibrils, an ability to bind thioflavin T and Congo red, and an extensive β-sheet structure.[9] The regions of the sequence that were found to promote aggregation of the TFE-denatured ensemble of AcP (residues 16−31 and 87−98) were found to be highly unfolded and flexible.[7,10]

Francesco Bemporad is a Ph.D. student at the Department of Biochemistry of the Università di Firenze, Firenze, Italy.

Giulia Calloni was a former Ph.D. student at the Department of Biochemistry of the Università di Firenze, Firenze, Italy. She is currently a postdoctoral researcher, supported with an EMBO fellowship, at the Department of Cellular Biochemistry of the Max-Planck-Institut für Biochemie, Martinsreid, Germany.

Silvia Campioni is a Ph.D. student at the Department of Biochemistry of the Università di Firenze, Firenze, Italy.

Georgia Plakoutsi was a former Ph.D. student at the Department of Biochemistry of the Università di Firenze, Firenze, Italy.

Niccolò Taddei is a Full Professor of Biochemistry at the Università di Firenze, Firenze, Italy. His research on protein folding and aggregation is carried out at the Department of Biochemistry of the Università di Firenze, Firenze, Italy.

Fabrizio Chiti received his B.S. degree at the Università di Firenze, Firenze, Italy, and his Ph.D. in chemistry at the Oxford University in U.K. His predoctoral research, performed under the supervision of Prof. C. M. Dobson, concerned protein-folding studies. His postdoctoral work, performed in the field of protein aggregation and amyloid formation, was carried out at the Università di Firenze for 2 years and at the University of Cambridge for 9 months. Starting from November 2002, Fabrizio Chiti is Associate Professor of Biochemistry at the Faculty of Medicine of the Università di Firenze. He is also a member of the EMBO Young Investigator Programme starting from 2005. The research activity of his group, established in the Department of Biochemistry of the Università di Firenze, currently focuses on protein folding, misfolding, and aggregation.

* To whom correspondence should be addressed. Telephone: +39-055-4598319. Fax: +39-055-4598905. E-mail: fabrizio.chiti@unifi.it.

10.1021/ar050067x CCC: $33.50
Published on Web 09/02/2006

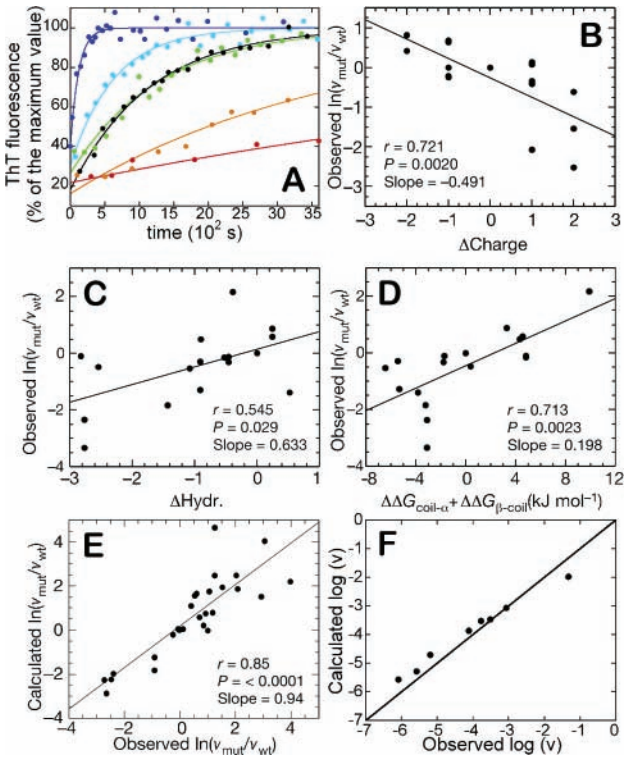

FIGURE 1. (A) Aggregation kinetics for a set of unfolded AcP variants obtained by recording the increase in ThT fluorescence. Wild type (black), A30G (blue), E29D (cyan), V20A (orange), Y98Q (red), and V47A (green). Solid lines represent the lines of best fit to single-exponential functions. Readapted with permission from ref 7. (B−D) Correlations between the change in the aggregation rate upon mutation for a set of unfolded AcP mutants and changes in the overall net charge (B), hydrophobicity (C), and free-energy variation associated with the conformational transition from α to β structure (D). Readapted with permission from ref 8. (E) Correlation between changes in the aggregation rate experimentally obtained for a set of mutations of short peptides or natively unfolded proteins, including amylin, amyloid β-peptide, τ, and α-synuclein, and the corresponding values calculated using the algorithm edited by Chiti and co-workers.[8] Readapted with permission from ref 8. (F) Correlation between the logarithm of the absolute aggregation rates (ν) obtained experimentally for a set of nonhomologous unfolded systems and the corresponding values calculated using the algorithm proposed by Vendruscolo and co-workers.[16] The solid line is not the best fit to a linear function but helps the reader to see the ideal agreement between experimental and theoretical values. The linear correlation coefficient (r) between predicted and observed rates is 0.91. Readapted with permission from ref 16.

Amino acid substitutions within such unstructured regions of the sequence can change the rate of aggregation of the denatured ensemble, as monitored with the ThT fluorescence assay[7] (Figure 1A). Changes in the aggregation rate upon mutation [$\ln(\nu_{mut}/\nu_{wt})$] correlate with changes in the hydrophobicity, in the propensity to undergo a conformational transition from α to β structure and in the net charge of the whole polypeptide chain[8] (parts B−D of Figure 1). This means that an increase in hydrophobicity and β-sheet propensity in key regions of the protein sequence results in an increased aggregation rate, whereas an increase in the overall net charge has the opposite effect as a result of the increased intermolecular repulsion. The observed correlations have been used to edit an algorithm expressing the change of the aggregation rate of an unfolded polypeptide chain following a given mutation [$\ln(\nu_{mut}/\nu_{wt})$] as a function of the changes in these three physicochemical factors caused by the same mutation. The resulting formula was found to reproduce the $\ln(\nu_{mut}/\nu_{wt})$ values for a number of mutations of unfolded systems other than AcP and involving, for example, Aβ, α-synuclein, and τ[8] (Figure 1E).

In another study, variants of the Aβ_{1-42} peptide, obtained by substituting Phe19 with all of the other 19 residues, were fused to the green fluorescent protein (GFP) and expressed in *Escherichia coli* cells.[11] The variants were shown to have aggregation propensities in the cells, determined by measuring the fluorescence emitted by the GFP, that correlate with both the changes of β-sheet propensity and hydrophobicity following mutation.[11] In particular, the best correlation was found with the predicted $\ln(\nu_{mut}/\nu_{wt})$ values determined using our formula.[11] In another study, a random mutagenesis was carried out on positions 41 and 42 of Aβ_{1-42} fused to GFP, showing a strong correlation between the aggregation propensity of the whole Aβ_{1-42} peptide, monitored both *in vitro* and *in vivo*, and the hydrophobicity and β-sheet propensity of the inserted residues.[12] Moreover, Aβ_{1-42} variants having hydrophilic or β-sheet breaker amino acids at these positions were less prone to aggregate than Aβ_{1-40}.[12]

Another related algorithm was edited that takes into account the change of polar and nonpolar water-accessible surface areas (negatively and positively linked to the aggregation rate, respectively), the change of the polar side-chain dipole moment, the change of propensity to form β-sheet structure, the change of charge, and the change of aromatic moieties number upon mutation.[13] This expression and its reciprocal form have been shown to predict the effect of a mutation on aggregation and disaggregation rates, respectively.[13] However, the role of aromatic residue interactions as a factor that facilitates aggregation is at present debated. Indeed, fragments of IAPP having a phenylalanine substituted by leucine were shown to retain a similar aggregation propensity.[14] In addition, it was shown that the deceleration of the aggregation rate of AcP following substitution of an aromatic residue with a nonaromatic one results mainly from the decrease of hydrophobicity and β-sheet propensity following mutation rather than from the elimination of the aromatic ring.[15]

Apart from the need to clarify the role of aromaticity in amyloid aggregation, these observations have represented the starting point of an increasing number of analyses. These have aimed, on the one hand, at improving our ability to rationalize and predict the change of the aggregation rate and/or propensity following mutation. On the other hand and perhaps more importantly, they have contributed to extend the idea that amyloid formation is governed by simple physicochemical factors to gain an understanding of the mechanism of amyloid aggregation in a broader manner.

The same three parameters mentioned above (hydrophobicity, propensity to form β-sheet structure, and net

charge) have been suggested as key determinants of the absolute aggregation rate of a fully unfolded protein, more exactly, the rate constant for the growth phase of fibril formation.[16] In this analysis, the hydrophobic pattern has been considered as a descriptor of β-sheet propensity,[16] following the demonstration that patterns of alternating hydrophilic and hydrophobic residues increase the aggregation propensity of a sequence because of their ability to form β-strands.[17] The aggregation rate constant of an unfolded system was expressed as a function of the hydrophobicity, β propensity, and net charge of the chain (intrinsic factors), as well as a function of extrinsic factors such as pH, ionic strength, and protein concentration.[16] Each of these factors was multiplied by a corresponding coefficient in the equation, to weigh its importance in determining the aggregation rate constant. The experimental values of the aggregation rate constant for a number of unfolded systems were used in a multiple regression fitting to determine the coefficients of all of the intrinsic and extrinsic factors in the formula. The resulting equation was shown to provide estimates of aggregation rate constants of unfolded polypeptide chains in good agreement with the data obtained experimentally[16] (Figure 1F).

Another recently proposed algorithm uses peptide solubility (as a parameter that is inversely proportional to aggregation propensity), polar and nonpolar water-accessible surface areas, secondary-structure propensities, and net charge of all of the amino acid residues in the sequence to determine the absolute aggregation rate of an unfolded polypeptide chain.[18] The authors also point out π stacking as a driving force in amyloid formation and use the total number of aromatic residues in the sequence as an additional factor in their equation.[18]

Identification of Regions of the Sequence That Promote Aggregation of Unfolded Polypeptide Chains

After the application of novel experimental approaches, it has become possible to identify experimentally the regions of the sequence that promote aggregation and form the cross-β core of the resulting fibrils. Using solid-state nuclear magnetic resonance (SS-NMR) and site-directed spin-labeling coupled to electron paramagnetic resonance (SDSL-EPR) analyses, it was possible, for example, to identify with reasonable accuracy the regions of the sequence that form the β core of the fibrils of the amyloid β peptide[19,20] (horizontal bars in Figure 2A). Similarly, the broad region spanning approximately residues 35−105 was found to be structured within the fibrils of α-synuclein, as determined using limited proteolysis,[21] SDSL-EPR,[22] and hydrogen/deuterium exchange.[23] Using SS-NMR spectroscopy, it was possible to identify, within such a broad region, some of the stretches that adopt a β-strand conformation relative to a more generic loop structure[24] (horizontal bars in Figure 2B).

After our increasing ability to gain information on the amyloid fibril structure at a molecular level, the question has arisen on whether one can rationalize why specific regions of an unstructured sequence are more aggregation-prone than others and on whether predictions can be made to identify such regions. An important extension of the concept that physicochemical factors govern the process of amyloid fibril formation of an unfolded polypeptide chain has resulted in the realization of a number of theoretical approaches in this direction.[18,25−27] By considering hydrophobicity, net charge, hydrophobic patterning, and secondary-structure propensities as parameters, a scale of aggregation propensities for the 20 naturally occurring amino acids was defined.[25]

An aggregation propensity profile was then obtained by calculating the average aggregation propensities over

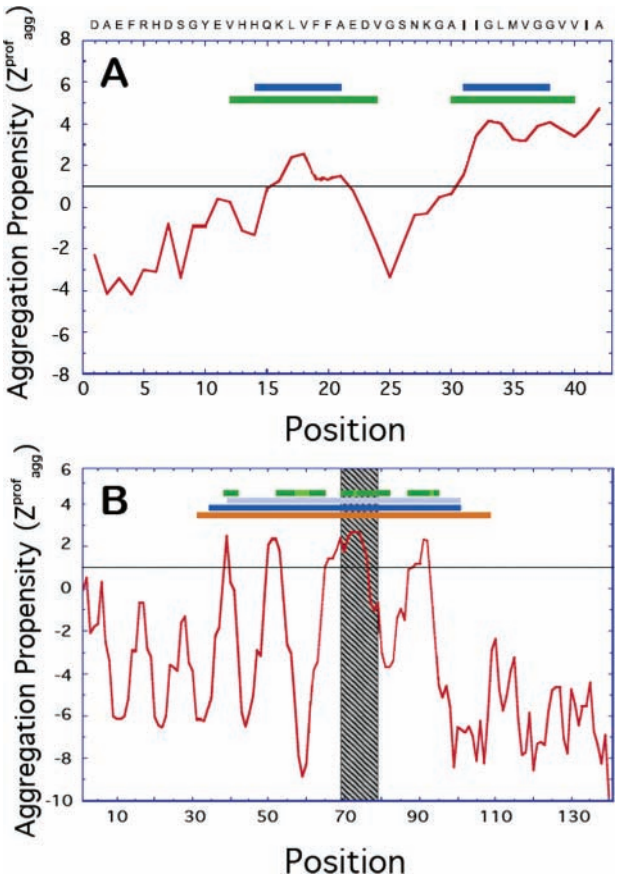

FIGURE 2. Amyloid aggregation profiles of Aβ_{1-42} (A) and α-synuclein (B) at pH 7.0 calculated as described.[25] A line at $Z^{prof}_{agg} = 1$ represents a threshold to identify the aggregation promoting regions. (A) Horizontal bars indicate regions of the sequence forming the core of the fibrils as determined with SS-NMR[19] (green) and SDSL-EPR[20] (blue). (B) Horizontal bars indicate the broad region of the sequence that appears to be structured in the fibrils from SDSL-EPR[22] (blue), hydrogen/deuterium exchange[23] (pale blue), and limited proteolysis[21] (orange). The regions adopting a β-strand conformation within such a region are indicated by green horizontal bars.[24] Because of experimental uncertainty, the boundaries of such strands are subject to some error.[24] The light green bars interrupting β strands indicate nonassigned amino acids or residues assigned to a disordered structure.[24] The reported data refer to structure b illustrated in Figure 5 of the cited paper. The hashed shadowed area refers to the high amyloidogenic 69−79 segment found to promote fibril formation of α-synuclein.[25] Readapted with permission from ref 25.

sliding windows of seven contiguous residues (attributing in each case the average value to the central residue of the window) and plotting the values as a function of the residue number (red traces in Figure 2). The choice of a sliding window of seven residues is a compromise between the need of lowering the noise in the profile resulting from shorter windows and the need of avoiding flat profiles resulting from larger windows. A very good agreement is observed between the peaks in the profile and the fragments that have been determined from experimental data to form the β core of the fibrils, as shown by the representative examples of Aβ and α-synuclein[25] (Figure 2). This shows that it is possible to predict the regions promoting aggregation of unstructured proteins from only the sequence.[25] Moreover, by considering all possible substitutions at each position, the regions of the sequence where mutations have the highest or lowest effect on the aggregation propensity were accurately calculated.[25] This is of particular interest for the development of mutational methods aimed at altering the aggregation propensity and mechanism of polypeptide chains.

Serrano and co-workers have developed a statistical mechanics algorithm (TANGO) that allows for the identification of aggregation-prone regions in proteins.[27] Assuming that the nucleating regions of an aggregate are fully buried, this algorithm calculates for each amino acid residue the percentage occupancy of the major competing conformational states, namely, α-helix, β-turn, β-sheet, the folded state (if any), and finally the β aggregate, on the basis of their relative energy content. For this computational task, TANGO takes into account different energetic terms (hydrophobicity, solvation energetics, electrostatic interactions, and hydrogen bonding) and the effects of solution parameters (pH, ionic strength, etc.). The highly aggregation-prone sequences provided by the algorithm correspond to segments of at least five consecutive residues populating the β-aggregated state at a more than 5% per residue. TANGO can also predict the change in aggregation propensity upon mutation.

As a follow-up of the study by Ventura and co-workers described above, the measured aggregation propensities of all of the 20 variants of Aβ_{1-42} obtained by substituting Phe19 with all possible residues were used to determine a scale of aggregation propensity for all 20 naturally occurring residues.[26] The profiles obtained using this empirical scale of aggregation propensity values identify the hot spots of aggregation for unfolded peptides and natively unfolded proteins (amylin, α-synuclein, and Aβ), in good agreement with the available experimental data.[26] The model can also predict the change in aggregation propensity upon mutation. The crystal structure of the Sup35p-derived peptide NNQQNY, assembled to form two adjacent parallel β-sheets, was used to model the structures and compute the related energies of all possible sequences of interest having six residues.[28] This approach was used to identify fibril-forming segments within proteins as those hexapeptides having an energy lower than a defined threshold.[28] Although the agreement between the predicted and experimentally determined segments is not yet satisfactory in this case, this structure-based approach is potentially interesting and will hopefully be able to provide important predictions following the refinement.

Other parameters have been proposed to play a fundamental role in determining the aggregation rate of fully unfolded proteins. A toy model has been recently proposed on the basis of a collision-encounter scheme.[29] This model assumes the existence of some key residues that promote nucleation and growth of the fibril. It shows that increasing the sequence distance between these residues decreases the rate of the process.[29] Moreover, the presence of flexible noninteracting tails slows down aggregation.[29] This decelerating effect is more marked when the region containing the key residues is positioned between two flanking regions with the same length, suggesting that the presence of aggregation-prone residues at the end of the sequence increases the rate of the process. While the conclusions of this toy model await experimental confirmation, these additional suggested parameters may represent important tools to improve the accuracy of the existing algorithms.

Determinants of the Aggregation of Partially Structured Proteins

Many of the natural proteins existing in living organisms adopt normally, under physiological conditions, partially structured states, with both unfolded and persistently folded domains. Well-known examples are all of the mammalian and fungal prion proteins thus far characterized. In addition, fully folded proteins can transiently adopt such conformational states, for example, during folding after biosynthesis, as a consequence of destabilizing mutations, trafficking, or stress.[6] It is widely accepted, for example, that the large majority of globular proteins need to unfold, at least partially, to polymerize into fibrillar aggregates.[2,6]

Under conditions that promote amyloid fibril formation of AcP in the presence of moderate concentrations of TFE, this α/β globular protein adopts initially a partially folded state in which several portions of the sequence populate native and non-native α-helices and the native β-hairpin formed by β-strands 2 and 3 is persistently formed.[7,9] A total of 33 mutants of AcP were produced, each having a single conservative substitution.[7] Only the mutations involving residues 16−31 and 87−98 were able to change the rate of aggregation of AcP from the partially folded state of the protein. Furthermore, dissection of the sequence of AcP into smaller fragments indicated that these two stretches are poorly soluble. It was therefore concluded that these two regions are the most relevant stretches of the sequence to promote the aggregation process of the protein.[7] While these two regions of the AcP sequence correspond effectively to peaks in the aggregation propensity profile (Figure 3A), other regions of the sequence appear to have a similar inherently high aggregation propensity. Using only sequence-based phys-

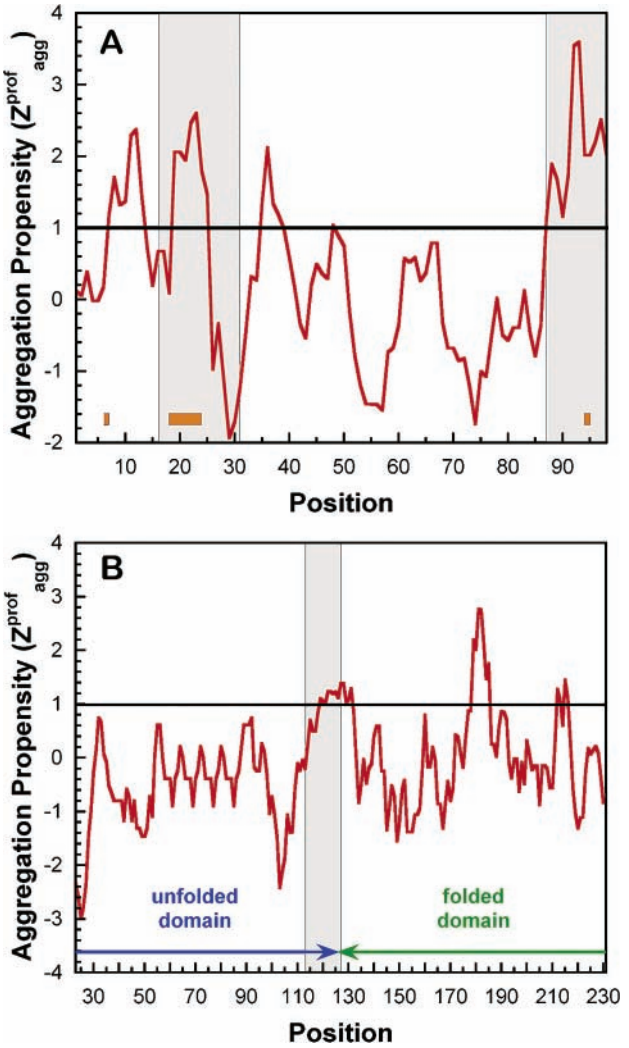

FIGURE 3. Aggregation propensity profiles of AcP at pH 5.5 (A) and human prion protein at pH 4.5 (B), calculated as described.[25] A line at $Z^{prof}_{agg} = 1$ is drawn to identify the threshold for aggregation-promoting regions. (A) Gray areas indicate the two aggregation-promoting regions as revealed by a protein-engineering approach.[7] The horizontal orange bars indicate regions susceptible to proteolysis in the amyloidogenic partially folded state.[10] (B) Gray shaded area indicates the aggregation-promoting region determined experimentally (see ref 31 and references therein). The blue and green arrows indicate the extensions of the unfolded and folded domains, respectively

icochemical factors, it does not therefore appear clear why such regions are not relevant for aggregation.

Clues to clarify this issue came from a detailed limited proteolysis analysis, an approach that allows for the specific residues or regions of the sequence that appear most exposed to the solvent and/or flexible to be identified.[10] When the monomeric amyloidogenic state of AcP was subjected to four proteases with different specificities, all proteolytic sites were found in three regions of the sequence: 6−7, 18−24, and 94−95 (Figure 3A). The two proteolytic regions 18−24 and 94−95 correspond to peaks in the aggregation propensity profile and, approximately, to the two regions of the sequence found experimentally to promote aggregation of the entire protein (Figure 3A). Importantly, no proteolytic sites were found within other regions of the sequence having a high intrinsic aggregation propensity, whereas the proteolyzed region 6−7 has an intrinsically low aggregation propensity (Figure 3A). In summary, our data on AcP suggest that the determinants of amyloid aggregation of partially structured proteins are both intrinsic sequence-based propensities and structural factors. Within such conformational states, aggregation will be promoted by regions of the sequence that have (i) an intrinsically high propensity to aggregate, as dictated by a high hydrophobicity and propensity to form β structure, but also (ii) a high degree of accessibility to the solvent and conformational flexibility.

Similar arguments hold for both mammalian and fungal prion proteins. Mammalian prion proteins are composed of an extended and flexible N-terminal segment (residues 23−125) and a well-defined globular domain (residues 126−231) comprising three α-helices and a short two-stranded antiparallel β-sheet (horizontal arrows in Figure 3B).[30] Experimental data obtained *in vitro* from both PrP fragments and deletion mutants show that one of the most important regions in promoting PrP aggregation spans approximately residues 113−127, whose sequence is conserved across all mammalian species (see ref 31 and references therein). Indeed, this region corresponds to a peak in the aggregation profile of the human prion protein (Figure 3B). The profile contains two other significant peaks, both located in the globular domain (Figure 3B). These regions are not directly involved in initiating aggregation because they form persistent and buried structures, but they may acquire an important role following destabilizing mutations.

The Sup35 prion protein from *Saccharomyces cerevisiae* can be divided into three distinct structural regions: the glutamine/asparagine-rich N domain (residues 1−123), the highly charged M domain (124−253), and a C-terminal GTP-binding domain (254−685). The structural core of Sup35 amyloid fibrils, found recently by a variety of experimental approaches to be composed by several strands contributed approximately by residues 21−121,[32] corresponds to a large region with an aggregation propensity that is in many segments higher than the threshold of $Z^{prof}_{agg} = 1$; the preceding (residues 1−20) and following (residues 158−253) regions of the NM sequence have low values of Z^{prof}_{agg}. The structured C domain contains many prominent peaks in the profile, but their roles in amyloid formation appear to be irrelevant as a result of the folded structure of these domains.

In conclusion, all of the data available thus far show that, similar to fully unfolded states, amyloid aggregation of partially unfolded proteins and proteins that contain both unfolded and folded domains in their native states is driven by specific regions of their sequence. In contrast to fully unfolded states, however, these regions must fulfill two requirements: (i) have a high intrinsic propensity to aggregate, as defined in the previous paragraph for unfolded systems, and (ii) have a high degree of chain flexibility. In agreement with this conclusion, it has been shown that the ease of fibril formation of apomyoglobin correlates with the degree of disorder observed in its

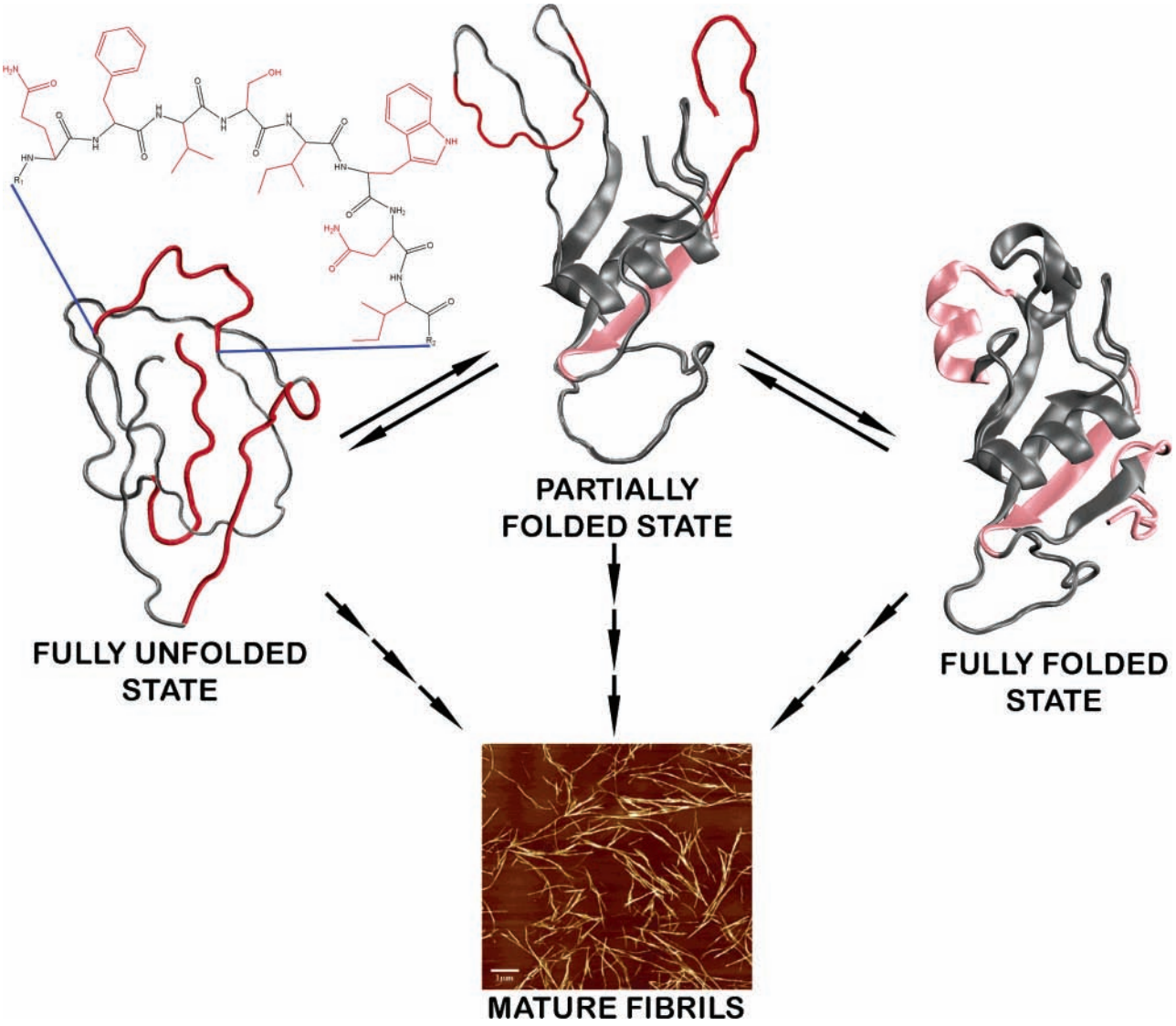

FIGURE 4. General mechanism for protein aggregation. Fully folded, partially folded, and fully unfolded states can all contribute to initiate aggregation. In fully unfolded states, aggregation is promoted by regions of the sequence with a high intrinsic propensity to aggregate (red). As shown, these fragments are rich in residues that are hydrophobic and/or with high β-sheet propensity. The aggregation of partially folded states is favored by the same type of fragments that remain unstructured (red), while those that become structured do not participate (pink). Nevertheless, the formation of amyloid protofibrils and fibrils can also start from a fully folded protein, albeit with minor propensity. The determinants of aggregation from the latter state still need to be clarified.

unfolded state.[33] An implemented algorithm to identify aggregation-promoting regions and applicable to fully unfolded as well as partially structured proteins will soon be published and can be run at http://www.zyggregator.com.

Fully Folded States Have a Small but Significant Propensity To Initiate Aggregation

The partial or complete unfolding of proteins that normally adopt compact folded structures causes their ability to aggregate to increase dramatically.[2,6] This results from the exposure and gained flexibility of regions of the sequence that are initially buried and therefore unavailable for intermolecular interactions. Two proteins investigated in our lab, AcP and the N-terminal domain from the *E. coli* HypF, actually form amyloid-like fibrils following partial unfolding.[34,35] Nevertheless, evidence is mounting that proteins retain some propensity to aggregate when folded into a compact globular structure. The acylphosphatase from *Sulfolobus solfataricus* (Sso AcP) converts rapidly into ordered aggregates with amyloid-like properties in 15−25% (v/v) TFE at pH 5.5 and 25 °C.[36,37] Under these conditions, before aggregation occurs, Sso AcP has a considerable enzymatic activity[36] and native-like far- and near-UV CD spectra[38] and folding is faster than unfolding.[36] Aggregation from this native-like conformation is also more rapid than unfolding, indicating that self-assembly of Sso AcP does not require a transition into a partially unfolded conformation.[36] Further analysis showed that native Sso AcP aggregates initially into species that do not possess extensive β-sheet structure, do not bind to the characteristic ThT and CR dyes, and are

enzymatically active, implying a native-like topology.[37] These oligomers reorganize directly into aggregates that have lost their enzymatic activity, contain extensive β structure, bind ThT and CR, and have a morphology and size, as revealed by transmission electron microscopy, typical of the spherical and chain-like amyloid protofibrils observed for Aβ, α-synuclein, and other systems.[37]

The acylphosphatase from *Drosophila melanogaster* (AcPDro2) forms amyloid-like fibrils under conditions in which the protein has initially a secondary structure, hydrodynamic diameter, catalytic activity, and packing around hydrophobic residues indistinguishable from those of the native state.[39] Importantly, in this case, the ΔG_{U-F} has the same value under native or aggregating conditions. Similar to Sso AcP, the kinetic analysis shows that AcPDro2 does not need to unfold to initiate aggregation.[39] The S6 protein from *Thermus thermophilus* adopts a quasi-native state at pH 2.0, 0.4 M NaCl, and 42 °C.[40] Under stirring, the protein grows into fibrils after several days. Kinetic analysis revealed that longer lag phases in aggregation correlate with faster unfolding rates in a number of variants, implying that the native-like state rather than an ensemble of highly fluctuating conformations, participates in the nucleation of the fibrillation process.[40] Bovine insulin forms, at pH 2.3 and 70 °C, initial aggregates characterized by a predominant native-like α-helical content, before forming β-sheet containing amyloid-like fibrils.[41]

Perhaps the most compelling evidence that "native" aggregation processes represent a real possibility for globular proteins, and a constant challenge for living organisms, is shown by the finding that natural all-β proteins have developed strategies during evolution to prevent their assembly through a direct interaction of folded units.[42] Aggregation processes that initiate from the folded conformations of globular proteins may acquire importance in pathology, considering that the majority of protein molecules spend most of their lifetime in a folded state and that protein deposition diseases have a slow onset even in the most acute cases.

Conclusions

Aggregation of a polypeptide chain may occur from a variety of conformational states, ranging from fully unfolded ensembles, lacking persistent secondary or tertiary structure, to fully folded compact structures (Figure 4). The assembly of proteins adopting initially an unfolded state is promoted by regions of the sequence that have an inherently high propensity to form β-structured aggregates, i.e., with a high hydrophobicity, high propensity to form β-sheet structure, and low net charge. These principles also seem to hold for proteins that adopt a distribution of conformations that deviates significantly from a totally random behavior but in which the elements of secondary structure or long-range interactions are not persistent, as in Aβ and α-synuclein.

Proteins that adopt, either transiently or enduringly, conformational states in which both unfolded stretches and regions or domains with persistent structure are present appear to aggregate via their unfolded regions, even when the folded parts present β-hairpins or other portions that are apparently ideal to initiate aggregation (Figure 4). Within the unfolded domains, aggregation is promoted, similarly to fully unfolded proteins, by regions of the sequence with an intrinsically high propensity to aggregate. Such unfolded portions initiate the aggregation process and form a substantial part of the β core in the fibrils. Some regions of the folded domains may also contribute, however, to the formation of the core, as in mammalian prions.

Finally, proteins in their fully folded state also retain a significant, albeit small, propensity to aggregate (Figure 4). It will be important to elucidate the factors governing aggregation processes from folded structures because these may have a considerable physiological relevance and represent an important issue in the chemistry of proteins. More efforts are in general required to identify those additional factors that govern protein aggregation processes in addition to those identified thus far and described here. Because amyloid fibril formation seems to be a generic feature of polypeptide chains and obey rules with general validity, we are optimistic that principles governing this apparently complex process will be within our reach as experimental data on the structure and mechanism of the formation of amyloid structures accumulate in the next few years.

Fabrizio Chiti's research is funded by EMBO (Young Investigator Programme 2005), the Italian MIUR (project FIRB RBNE03PX83), and the European Union (project number HPRN-CT-2002-00241). All authors contributed equally to the work and are listed in alphabetical order (except the corresponding author).

References

(1) Stefani, M.; Dobson, C. M. Protein aggregation and aggregate toxicity: New insights into protein folding, misfolding diseases and biological evolution. *J. Mol. Med.* **2003**, *81*, 678−699.
(2) Uversky, V. N.; Fink, A. L. Conformational constraints for amyloid fibrillation: The importance of being unfolded. *Biochim. Biophys. Acta* **2004**, *1698*, 131−153.
(3) Chiti, F.; Dobson, C. M. Protein misfolding and diseases. *Ann. Rev. Biochem.* **2006**, *75*, 333−366.
(4) Chapman, M. R.; Robinson, L. S.; Pinkner, J. S.; Roth, R.; Heuser, J.; Hammar, M.; Normark, S.; Hultgren, S. J. Role of *Escherichia coli* curli operons in directing amyloid fiber formation. *Science* **2002**, *295*, 851−855.
(5) Chien, P.; Weissman, J. S.; DePace, A. H. Emerging principles of conformation-based prion inheritance. *Annu. Rev. Biochem.* **2004**, *73*, 617−656.
(6) Dobson, C. M. Protein folding and misfolding. *Nature* **2003**, *426*, 884−890.
(7) Chiti, F.; Taddei, N.; Baroni, F.; Capanni, C.; Stefani, M.; Ramponi, G.; Dobson, C. M. Kinetic partitioning of protein folding and aggregation. *Nat. Struct. Biol.* **2002**, *9*, 137−143.
(8) Chiti, F.; Stefani, M.; Taddei, N.; Ramponi, G.; Dobson, C. M. Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* **2003**, *424*, 805−808.
(9) Chiti, F.; Webster, P.; Taddei, N.; Clark, A.; Stefani, M.; Ramponi, G.; Dobson, C. M. Designing conditions for in vitro formation of amyloid protofilaments and fibrils. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 3590−3594.

(10) Monti, M.; Garolla di Bard, B. L.; Calloni, G.; Chiti, F., Amoresano, A.; Ramponi, G.; Pucci, P. The regions of the sequence most exposed to the solvent within the amyloidogenic state of a protein initiate the aggregation process. *J. Mol. Biol.* **2004**, *336*, 253−262.
(11) de Groot, N. S.; Aviles, F. X.; Vendrell, J.; Ventura, S. Mutagenesis of the central hydrophobic cluster in Aβ42 Alzheimer's peptide. *FEBS J.* **2006**, *273*, 658−668.
(12) Kim, W.; Hecht, M. H. Sequence determinants of enhanced amyloidogenicity of Alzheimer Aβ_{42} peptide relative to Aβ_{40}. *J. Biol. Chem.* **2005**, *41*, 35069−35076.
(13) Tartaglia, G. G.; Cavalli, A.; Pellarin, R.; Caflisch, A. The role of aromaticity, exposed surface, and dipole moment in determining protein aggregation rates. *Protein Sci.* **2004**, *13*, 1939−1941.
(14) Tracz, S. M.; Abedini, A.; Driscoll, M.; Raleigh, D. P. Role of aromatic interactions in amyloid formation by peptides derived from human amylin. *Biochemistry* **2004**, *43*, 15901−15908.
(15) Bemporad, F.; Taddei, N.; Stefani, M.; Chiti, F. Assessing the role of aromatic residues in the amyloid aggregation of human muscle acylphosphatase. *Protein Sci.* **2006**, *15*, 862−870.
(16) DuBay, K. F.; Pawar, A. P.; Chiti, F.; Zurdo, J.; Dobson, C. M.; Vendruscolo, M. Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains. *J. Mol. Biol.* **2004**, *341*, 1317−1326.
(17) West, M. W.; Wang, W.; Patterson, J.; Mancias, J. D.; Beasley, J. R.; Hecht, M. De novo amyloid proteins from designed combinatorial libraries. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11211−11216.
(18) Tartaglia, G.; Cavalli, A.; Pellarin, P.; Caflisch, A. Prediction of aggregation rate and aggregation-prone segments in polypeptide sequences. *Protein. Sci.* **2005**, *14*, 2723−2734.
(19) Petkova, A. T.; Ishii, Y.; Balbach, J. J.; Antzukin, O. N.; Leapman, R. D.; Delaglio, F.; Tycko, R. A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 16742−16747.
(20) Torok, M.; Milton, S.; Kayed, R.; Wu, P.; McIntire, T.; Glabe, C. G.; Langen, R. Structural and dynamic features of Alzheimer's Aβ peptide in amyloid fibrils studied by site-directed spin labeling. *J. Biol. Chem.* **2002**, *277*, 40810−40815.
(21) Miake, H.; Mizusawa, H.; Iwatsubo, T.; Hasegawa, M. Biochemical characterization of the core structure of α-synuclein filaments. *J. Biol. Chem.* **2002**, *277*, 19213−19219.
(22) Der-Sarkissian, A.; Jao, C. C.; Chen, J.; Langen, R. Structural organization of α-synuclein fibrils studied by site-directed spin labeling. *J. Biol. Chem.* **2003**, *278*, 37530−37535.
(23) Del Mar, C.; Greenbaum, E. A.; Mayne, L.; Englander, S. W.; Woods, V. L., Jr. Structure and properties of α-synuclein and other amyloids determined at the amino acid level. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15477−15482.
(24) Heise, H.; Hoyer, W.; Becker, S.; Andronesi, O. C.; Riedel, D.; Baldus, M. Molecular-level secondary structure, polymorphism, and dynamics of full-length α-synuclein fibrils studied by solid-state NMR. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15871−15876.
(25) Pawar, A. P.; Dubay, K. F.; Zurdo, J.; Chiti, F.; Vendruscolo, M.; Dobson, C. M. Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases. *J. Mol. Biol.* **2005**, *350*, 379−392.
(26) de Groot, N. S.; Pallarés, I.; Avilés, F. X.; Vendrell, J.; Ventura, S. Prediction of "hot spots" of aggregation in disease-linked polypeptides. *BMC Struct. Biol.* **2005**, *5*, 18.
(27) Fernandez-Escamilla, A. M.; Rousseau, F.; Schymkowitz, J.; Serrano, L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.* **2004**, *22*, 1302−1306.
(28) Thompson, M. J.; Sievers, S. A.; Karanicolas, J.; Ivanova, M. J.; Baker, D.; Eisenberg, D. The 3D profile method for identifying fibril forming segments of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 4074−4078.
(29) Hall, D.; Hirota, N.; Dobson, C. M. A toy model for predicting the rate of amyloid formation from unfolded protein. *J. Mol. Biol.* **2005**, *195*, 195−205.
(30) Riesner, D. Biochemistry and structure of PrP^C and PrP^{Sc}. *Br. Med. Bull.* **2003**, *66*, 21−33.
(31) Hölscher, C.; Delius, H.; Bürkle, A. Overexpression of nonconvertible PrP^C Δ114−121 in scrapie-infected mouse neuroblastoma cells leads to trans-dominant inhibition of wild-type PrP^{Sc} accumulation. *J. Virol.* **1998**, *72*, 1153−1159.
(32) Krishnan, R.; Lindquist, S. L. Structural insights into a yeast prion illuminate nucleation and strain diversity. *Nature* **2005**, *435*, 765−772.
(33) Fändrich, M.; Forge, V.; Buder, K.; Kittler, M.; Dobson, C. M.; Diekmann, S. Myoglobin forms amyloid fibrils by association of unfolded polypeptide segments. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 15463−15468.
(34) Chiti, F.; Taddei, N.; Bucciantini, M.; White, P.; Ramponi, G.; Dobson, C. M. Mutational analysis of the propensity for amyloid formation by a globular protein. *EMBO J.* **2000**, *19*, 1441−1449.
(35) Marcon, G.; Plakoutsi, G.; Canale, C.; Relini, A.; Taddei, N.; Dobson, C. M.; Ramponi, G.; Chiti, F. Amyloid formation from HypF-N under conditions in which the protein is initially in its native state. *J. Mol. Biol.* **2005**, *347*, 323−335.
(36) Plakoutsi, G.; Taddei, N.; Stefani, M.; Chiti, F. Aggregation of the acylphosphatase from *Sulfolobus solfataricus*. *J. Biol. Chem.* **2004**, *279*, 14111−14119.
(37) Plakoutsi, G.; Bemporad, F.; Calamai, M.; Taddei, N.; Dobson C. M.; Chiti, F. Evidence for a mechanism of amyloid formation involving molecular reorganisation within precursor aggregates. *J. Mol. Biol.* **2005**, *351*, 910−922.
(38) Plakoutsi, G.; Bemporad, F.; Monti, M.; Pagnozzi, D.; Pucci, P.; Chiti, F. Exploring the mechanism of formation of native-like and precursor amyloid oligomers for the native acylphosphatase from *Sulfolobus solfataricus*. *Structure* **2006**, *14*, 993−1001.
(39) Soldi, G.; Bemporad, F.; Torrassa, S.; Relini, A.; Ramazzotti, M.; Taddei, N.; Chiti, F. Amyloid formation of a protein in the absence of unfolding and destabilisation of the native state. *Biophys. J.* **2005**, *89*, 4234−4244.
(40) Pedersen, J. S.; Christensen, G.; Otzen, D. E. Modulation of S6 fibrillation by unfolding rates and gatekeeper residues. *J. Mol. Biol.* **2004**, *341*, 575−588.
(41) Bouchard, M.; Zurdo, J.; Nettleton, E. J.; Dobson, C. M.; Robinson, C. V. Formation of insulin amyloid fibrils followed by FTIR simultaneously with CD and electron microscopy. *Protein Sci.* **2000**, *9*, 1960−1967.
(42) Richardson, J. S.; Richardson, D. C. Natural β-sheet proteins use negative design to avoid edge to edge aggregation. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *99*, 2754−2759.

AR050067X