ELSEVIER

Computational models for the prediction of polypeptide aggregation propensity

Amedeo Caflisch

In amyloid fibrils, β-strand conformations of polypeptide chains, or segments thereof, are perpendicular to the fibril axis, but knowledge of their three dimensional structure at atomic level of detail is scarce. Two types of computational approaches have been developed recently for investigating the aggregation propensity of peptides and proteins and identifying the segments most prone to form fibrils (hot spots). The physicochemical properties of the natural amino acids (e.g. β-propensity, hydrophobicity, aromatic content and charge) have been used to derive phenomenological models able to predict changes in aggregation rate upon mutation, as well as absolute rates and hot spots. Applications of these models to entire proteomes have provided evidence that intrinsically disordered proteins are less amyloidogenic than globular proteins. In the second type of approach, amyloidogenic polypeptides have been decomposed into overlapping segments, and atomistic simulations of three or more copies of each segment have been performed to obtain insights into aggregation propensity and structural details of the ordered aggregates (e.g. turn regions).

Addresses
Department of Biochemistry, University of Zurich, CH-8057 Zurich, Switzerland

Corresponding author: Caflisch, Amedeo (caflisch@bioc.unizh.ch)

Current Opinion in Chemical Biology 2006, **10**:437–444

This review comes from a themed issue on
Analytical techniques
Edited by Saul Tendler and Ehud Gazit

Available online 1st August 2006

1367-5931/$ – see front matter

DOI 10.1016/j.cbpa.2006.07.009

Introduction

β-sheet structures, which are, together with α-helices, one of the most common regular conformations of the polypeptide chain in folded proteins, have an intrinsic tendency to favor the formation of amyloid fibrils. Amyloid fibrils consist of ordered β-aggregates of the same protein in which the (poly)peptide chains are in an extended β-conformation, with their backbone perpendicular to the axis of the fibril [1]. Such β-aggregation is typical of amyloid diseases, which include Alzheimer's, Parkinson's and type II diabetes, as well as the spongiform encephalopathies. There is currently no effective treatment against these progressive disorders, most of which affect the brain in a devastating way. Therefore, it is of fundamental medical interest to understand the mechanisms of fibrillogenesis with the ultimate goal of determining the relative toxicity of (soluble) oligomers, protofibrils and mature fibrils, and designing small molecules that interfere with, and ideally inhibit, the formation of the toxic species. Two important and related issues regarding amyloid fibril formation are the specificity with which the amino acid sequence determines β-aggregation propensity and the atomic details of the fibril structure. The few atomic-level structures of amyloid fibrils that are available have been attained by X-ray microcrystallography data [2], by combining a range of biophysical techniques including fluorescence and NMR spectroscopy [3], or by applying quenched hydrogen–deuterium exchange NMR together with pairwise mutagenesis [4•]. Because of the difficulties in obtaining detailed structural information by X-ray crystallography or solution NMR spectroscopy, computational approaches are also needed. These approaches could help determine the short segments of amyloid-like polypeptides that share the same biophysical properties of the full-length proteins and identify those elements that are essential for the formation of amyloid fibrils. Recent review articles on computational studies of fibril formation have focused on the design of model systems (e.g. short peptides) for amyloid formation [5•] and on approaches based on molecular dynamics simulations [6–8].

The focus of this review is on recent approaches to predict aggregation rates and amyloidogenic segments of polypeptide sequences. The emphasis is on phenomenological models that use the physicochemical properties of the sidechains, and computational methods based on atomistic descriptions of β-aggregation. These approaches have provided interesting insights into the complex process of ordered aggregation. Most of these methods and their applications were published in 2004–2006 and are still subject of intense investigations.

Phenomenological models based on physicochemical properties

Chiti and coworkers [9] have been the first to quantitatively analyze the effects of mutations on polypeptide aggregation rates. They have observed that hydrophobicity, charge and propensity to convert from an α-helical to a β-sheet conformation influence the aggregation rate under conditions at which the considered proteins are mainly unstructured. On the assumption that these three factors are independent of each other, they have proposed an empirical model to predict the effect of a single-point

mutation on the aggregation rate. Their linear three-parameter model approximates the natural logarithm of the ratio of the aggregation rate constants of the mutant (rate_mutant) and wild-type (rate_wt) peptide or protein (ln[rate_mutant/rate_wt]). When lag (i.e. nucleation) and fibril-elongation phases were distinguishable in the kinetic profiles of aggregation, only the elongation rate was considered. The three parameters were fitted using a set of 28 mutations of human muscle acylphosphatase and a correlation coefficient of 0.756 and slope of 0.96 were obtained. The cross-validation, on a set of 27 single-point mutants of amyloidogenic peptides not used to fit the model, yielded values of 0.85 and 0.94 for the correlation coefficient and slope, respectively [9]. Although the predictive ability is very high, it is important to note that the 12 mutants with different net charges were predicted with an average error of 1.08 and a maximal error of 3.39 (for the R406W mutant of tau protein), whereas the average and maximal experimental errors were 0.7 and 1.5, respectively. Such relatively large deviations in predicting ln(rate_mutant/rate_wt) might originate from the fact that charge and hydrophobicity (i.e. two of the three properties used in the linear model of Chiti and coworkers) are not completely independent from each other. In fact, the predictive power is much higher for the 15 mutations not involving charged residues, with average and maximal error of 0.35 and 1.03 (for the S20G mutant of the human islet amyloid polypeptide), respectively. Additional evidence for a systematic error related to charged sidechains, in particular arginine, can be found in a recent paper by Chiti and coworkers [10] where the decrease in aggregation rate for the acylphosphatase mutants F22R, Y91R and Y98R is overestimated by 0.9, 1.2 and 0.5 logarithmic units, respectively. However, a very high correlation was observed recently between relative aggregation propensity, predicted by the model of Chiti and coworkers [9], and *in vivo* fluorescence of different single-point mutants of the 42-residue human beta-amyloid peptide (Aβ42) in a green fluorescent protein fusion [11].

To predict absolute aggregation rates, DuBay and coworkers have extended the three-parameter equation [9] into a seven-parameter formula that includes intrinsic properties of the polypeptide chain and extrinsic factors related to the environment (e.g. peptide concentration, pH value and ionic strength of the solution) [12]. The intrinsic properties are hydrophobicity, charge and the patterns of alternating hydrophobic–hydrophilic residues [13], whereas the propensity to convert an α-helical to a β-sheet conformation (one of the three factors of [9]) is not taken into account in the seven-parameter formula. To fit the seven parameters, they have used experimental data from 59 mutants of acylphosphatase, as well as data from 18 mutants of amyloidogenic peptides (including Aβ40, Aβ42, human islet amyloid polypeptide and prion protein peptide PrP106-126) and proteins. Despite the large number of data points used in the fitting (77 experimentally measured elongation rates), the multiplicative parameters for pH and concentration could not be fitted in a robust way because of the limited size of the database, with most experiments performed at a single value of pH (5.5) and concentration (0.04 mM). As explicitly mentioned by the authors, the main limitation of the seven-parameter model is that all residues in the polypeptide sequence have the same relative importance, which is not consistent with a plethora of experimental observations [4•,14,15•] and simulation results [16•].

By focusing on the physicochemical properties that determine ordered aggregation, Tartaglia and coworkers [17] have developed a phenomenological model without free parameters to predict changes in elongation rate upon mutation. The physicochemical properties used are the change in β-propensity upon mutation [18,19], the change in number of aromatic residues [20–22] and the change in total charge. Furthermore, the ratio of accessible surface area is taken into account if the wild-type and mutant sidechains are both polar or both apolar, whereas the dipole moment of the polar sidechain is used in the case of apolar to polar (or polar to apolar) mutation. The model of Tartaglia and coworkers [17] has a more complicated functional form than the linear three-parameter model of Chiti and coworkers [9] but has the advantage that it does not require any fitting, being devoid of free parameters. Furthermore, it does not suffer of the partial redundancy of charge and hydrophobicity, which is present in the empirical models discussed above [9,12]. Interestingly, the model of Tartaglia reproduced the relative aggregation propensity of a set of 26 heptapeptide sequences, which were predicted to favor an in-register parallel β-sheet arrangement [17] by using molecular dynamics simulations of three heptapeptides in a box [23,24].

Recently, Tartaglia and coworkers [25•] have further developed their model to predict absolute elongation rates and identify aggregation-prone segments. The extended model distinguishes between parallel and antiparallel β-sheet arrangement according to the preponderance of apolar versus polar residues. A positional effect taken into account, but rewarded only in the case of antiparallel arrangement, is the number of sidechain pairs with opposite charges in symmetric position with respect to the center of the considered segment, such as Lys and Glu in the Aβ segment KLVFFAE. An essential element in the derivation of the model has been the analysis of a large pool of β-aggregating peptide sequences designed by a computational approach based on implicit solvent molecular dynamics and genetic algorithm optimization in sequence space (Tartaglia and Caflisch, unpublished results). Although some of the physicochemical properties in the model of Tartaglia and coworkers [17,25•] are similar to those used in the models of Chiti and coworkers [9,12], it is important to distinguish that the former is a

phenomenological model without parameters, whereas the latter is an empirical approach based on parameter optimization for a multiterm equation. Comparable results have been reported on the test set of experimentally measured aggregation rates used to develop these models [12,25•]. Although there is no in-depth comparison of these two models on a large set of experimental data not used to derive them, the seven-parameter approach is not expected to have a better predictive ability. One reason is that positional effects are explicitly taken into account in the model of Tartaglia and coworkers [25•], whereas the multiparameter approach is mainly based on amino acid composition [12]. Furthermore, the multiparameter approach cannot be used to identify β-aggregating segments, as explicitly mentioned by the authors [12]. Recently, the model of Chiti, Vendruscolo and coworkers [12] has been modified for the prediction of aggregation-prone segments of proteins and validated on Aβ42, α-synuclein and the tau protein [26]. The strongest contribution in their new model is hydrophobicity (see Table 2 of [26]) so that the approach is likely to overpredict segments with many apolar residues and miss some polar amyloidogenic segments, such as NNQQNY [2] and the polyglutamines.

Serrano and coworkers [27] have developed an approach based on secondary structure propensity and estimation of desolvation penalty (TANGO) to predict β-aggregating regions of a protein sequence as well as mutational effects. In contrast to the models discussed above, TANGO takes into account the native state stability by using the FOLD-X force field [28]. For each residue in a polypeptide chain, TANGO evaluates the percent occupancy of the β-aggregation conformation. A polypeptide is considered to have β-aggregation tendency if it contains a segment of at least five consecutive residues with a β-aggregation occupancy higher than 5%. TANGO is based on the assumption that the probability of finding more than two ordered segments in the same polypeptide is negligible. The authors report that TANGO is useful for quantitative comparison of relative aggregation propensities of mutants of a common sequence. However, it is not possible to calculate absolute rates of aggregation with TANGO, which provides only a qualitative comparison between peptides or proteins differing significantly in sequence.

Applications of phenomenological models

Proteome analysis of ordered aggregation

Serrano and coworkers [29•] have used TANGO to analyze the β-aggregation propensity of a set of nonredundant globular proteins with an upper limit of 40% sequence identity. They have provided evidence that the β-aggregation tendency of all-α, all-β and mixed α/β globular proteins as well as membrane-associated proteins is fairly similar. Furthermore, in a set of 296 intrinsically disordered proteins TANGO identified that aggregation-prone segments were three times fewer than in globular proteins, indicating that the formation of a globular protein comes at the cost of a higher β-aggregation propensity [29•]. The same authors have recently used TANGO to investigate the aggregation propensity of proteins in 28 complete proteomes spanning all kingdoms of life [30]. They have shown that evolutionary pressure minimizes the amount of strongly aggregating sequences. Moreover, evolution has favored capping of amyloidogenic segments by arginine, lysine and proline to improve selectivity for chaperone binding [30].

The model developed to predict absolute rates and aggregation-prone segments [25•] has been applied to the proteomes of nine eukaryotes [31]. One interesting finding is that proteomes of higher and more long-lived eukaryotes contain fewer sequences with high β-aggregation propensity and are accrued in proteins with low β-aggregation propensity. It was also observed that, compared with random proteomes (obtained by shuffling residues but keeping constant the number and length of proteins as well as global amino acid composition), natural proteomes are enriched in proteins with low β-aggregation potential as well as proteins with high β-aggregation potential. Such polarization is a consequence of a dual evolutive requirement: the formation of a folded protein structure with thermodynamically stable native state goes at the cost of higher β-aggregation propensity (as suggested previously [29•]), whereas intrinsically disordered proteins must have low β-aggregation tendency [31]. A recent application of the same phenomenological model to the proteome of the yeast *Saccharomyces cerevisiae* indicates that the evolutionary pressure has reduced the number of proteins with β-aggregation tendency in cellular compartments such as the nucleolus, characterized by a high concentration of unfolded molecules (Tartaglia and Caflisch, unpublished data).

Design of polypeptide mutants with reduced aggregation propensity

One challenging application of the phenomenological models is the prediction of a small mutation, ideally single-point, resulting in a large change in aggregation rate. The empirical model of Chiti and coworkers [9] was used recently to suggest variants of the 32-residue peptide hormone calcitonin to reduce its aggregation propensity [32]. In fact, calcitonin is used in the treatment of osteoporosis, Paget's disease, hypercalcemia and musculoskeletal pain, but its tendency to aggregate is a serious problem during production, storage and administration. The change in aggregation rate of more than 600 variants, each with one to six mutated sidechains, was evaluated with the three-parameter model [9], and the two mutated calcitonin with the slowest predicted rate were tested *in vitro*. Aggregation kinetics monitored by turbidity measurements confirmed that their stability in solution is higher than the wild-type human calcitonin. These two

variants have five and six mutations, of which three and four are arginine, respectively, and the remaining two threonine and serine. Hence, the reduction of aggregation tendency (which was largely overestimated by the three-parameter model, as reported in Table 1 of [32]) is not surprising because human calcitonin is amidated at the C-terminus and has only two oppositely charged sidechains at neutral pH (D15 and K18). Therefore, the increase in positive charge of the two variants due to the additional arginine sidechains is expected to improve solubility.

Computational methods at atomic resolution

Despite much higher computational demand with respect to the phenomenological models reviewed above, the approaches based on molecular dynamics provide a structural interpretation of the β-aggregation profile, which is very useful to rationalize the sequence dependence and predict mutational effects on amyloid aggregation. For instance, a significant decrease in the aggregation propensity of the N47S,N48S double-mutant of the 94-residue N-terminal domain of the yeast prion-like protein Ure2p was predicted with the molecular dynamics approach of Cecchini and coworkers [16•] and validated *in vitro* by monitoring the kinetics of aggregation using the thioflavin T binding assay.

The methods reviewed in this section are based on atomistic descriptions of the polypeptide chain. As such, they cannot be used for proteomic analysis but they provide detailed structural information, which is not possible to obtain with the phenomenological models.

Methods based on templates from experimental structures

Lopez de la Paz and coworkers [33] have been the first to design sequences that fit to a structural template of an ordered β-sheet aggregate. The sequence design was carried out by using a rigid template of six hexapeptides arranged in an antiparallel β-sheet, corresponding to the backbone of the large single-layer β-sheet of the outer surface protein A. Their design approach, which is based on a force field and implicit solvation model, was originally developed to improve the thermodynamic stability of β-sheet peptides [34], and specifically to try to solve the inverse structure prediction problem of finding sequences that fold into a given structure. They designed *de novo* the amyloid peptide STVIIE [33], which was then used to perform an exhaustive mutagenesis analysis [14]. Individual residues of STVIIE were systematically replaced with all natural amino acids except Cys. The positional scanning of the STVIIE hexapeptide revealed that amyloid fibril formation is strongly dependent on the position of the mutation. From the mutagenesis experiments, Lopez de la Paz and Serrano derived a sequence pattern that can be used to scan polypeptide sequences for six-residue segments that are potentially amyloidogenic. Their approach and pattern rely on the assumption that residue preferences at a given position are independent of the residue types at other positions. This assumption might result in a significant number of false positives. As an example, at neutral pH the hexapeptides with sequence E/D-E/D-V/L/S/W/F/N/Q-I/L/T/Y/W/F/N-F/I/Y-E/D are predicted to be amyloidogenic by the pattern of Lopez de la Paz and Serrano, but many (or even most) of these 1176 hexapeptides are likely not to aggregate because of electrostatic repulsion between the negative charges. Moreover, the pattern fails to recognize known fibril-forming hexapeptides, such as the segment NFGAIL in the human islet amyloid polypeptide and NNQQNY [2]. Despite these limitations, the authors have provided evidence that highly amyloidogenic motifs matching their pattern are underrepresented in natural proteins and appear to be surrounded by residues that inhibit their aggregating tendency, such as proline and charged sidechains [14].

Recently, Eisenberg, Baker and coworkers [35] have used the X-ray microcrystal structure of the cross-β spine formed by the peptide NNQQNY [2] as a template to identify amyloidogenic segments. Starting from the coordinates of the atoms in the microcrystal, a set of 2511 near-native structures (which the authors call 3D profile) was generated by translational shifting of a four-stranded parallel β-sheet with respect to a three-stranded parallel β-sheet with antiparallel arrangement of the two β-sheets. Each six-residue segment of the polypeptide of interest was threaded onto each of the 2511 near-native templates, and the energetic fit was evaluated using the ROSETTADESIGN program [36]. The energy was evaluated for an infinite 1D periodic system. In this regard, it has to be noted that the intra- and inter-β-sheet arrangement of the peptides in the microcrystal might not be representative of the (proto)fibril. For instance, being derived from a microcrystal with 1D periodicity, the X-ray structure of NNQQNY [2] does not show either the β-sheet twist or the common fibril twist derived from synchrotron X-ray diffraction patterns [37]. Another main assumption of the 3D profile method is the in-register parallel β-sheet structure, as observed in the microcrystal [2]. In-register parallel arrangements favor the interactions of hydrophobic and aromatic sidechains. Although a preference for parallel β-sheet aggregates is expected for polypeptide sequences with few charged sidechains, short stretches with sidechains of opposite charge at the termini might prefer the antiparallel arrangement [14,25•,38].

Yoon and Welsh [39] have developed a structure-based approach for detecting β-aggregation propensity of a protein segment conditioned on the number of tertiary contacts. Using a sliding seven-residue window, segments with a strong β-sheet tendency in a tightly packed environment (i.e. with a high number of tertiary contacts) were suggested to be local mediator of fibril formation. They

have investigated 2358 nonhomologous protein domains and provided evidence that most proteins contain segments with significant hidden β-strand propensity. Recently, the same authors have developed a procedure based on an artificial neural network for the prediction of contact-dependent secondary structure propensity [40]. An analysis of 1930 nonhomologous protein domains has revealed that the α-helix and β-strand share similar sequence context, and that the number of tertiary contacts is an important determinant of the native secondary structure.

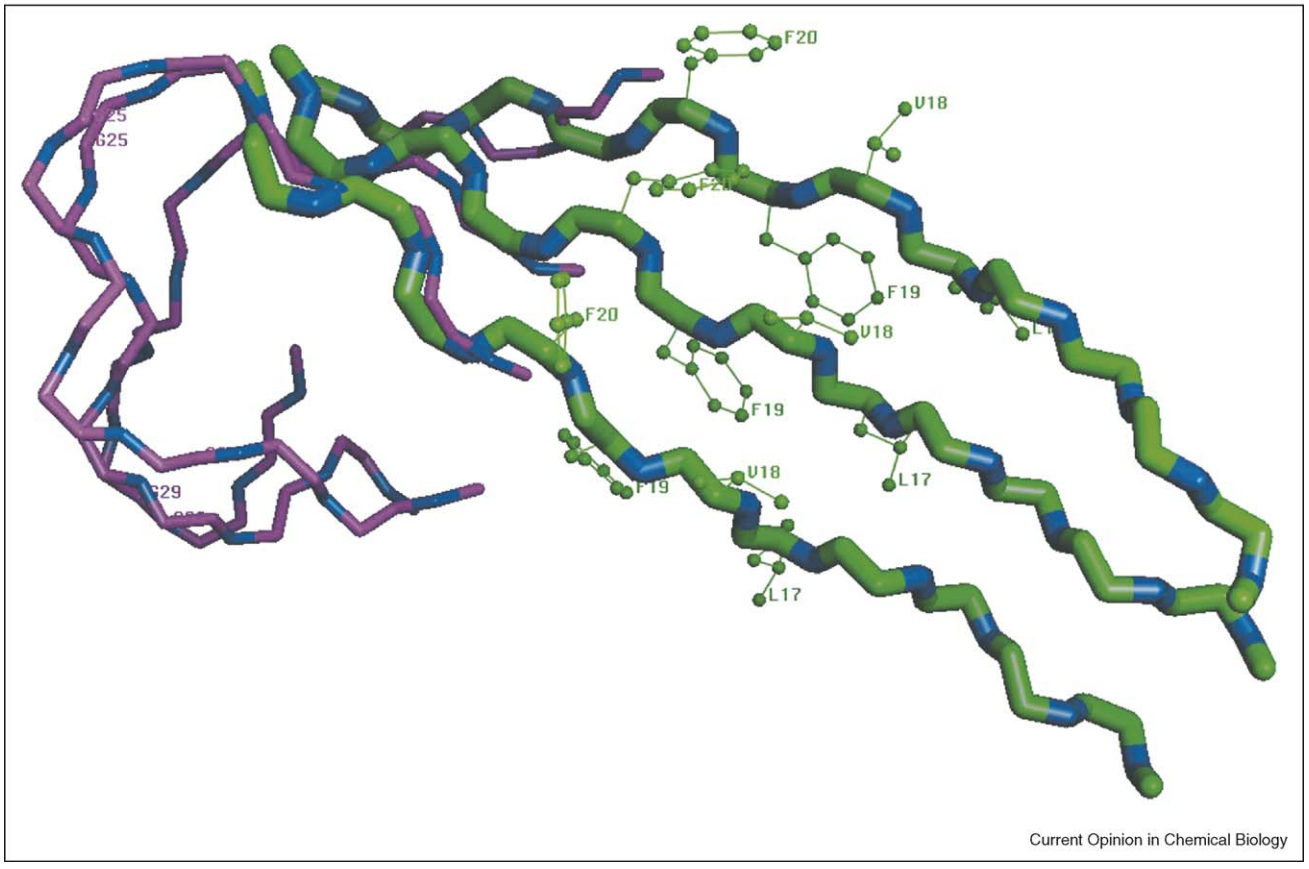
Approaches based on molecular dynamics

There is a review on molecular dynamics simulations of aggregation by Ruth Nussinov in this issue of *Current Opinion in Chemical Biology*. Therefore, this subsection focuses on two papers in which multiscale simulation techniques were used to identify amyloidogenic stretches and obtain the β-aggregation profile of polypeptides that are known to aggregate *in vitro*.

Dokholyan and coworkers [41•] have investigated the ordered aggregation propensity along the sequence of the enzyme Cu, Zn superoxide dismutase (SOD1), several mutations of which are linked to the familial form of amyotrophic lateral sclerosis. In its native state, SOD1 is homodimeric and each monomer adopts the Greek-key fold, a β-barrel consisting of two four-stranded β-sheets connected by two crossover loops. They have decomposed the SOD1 sequence into overlapping heptapeptides and performed a large number of explicit water molecular dynamics simulations (each of 0.5 ns) of monomeric, dimeric and tetrameric segments. Each segment was *N*-acetylated and *C*-amidated to reproduce the original context in the full-length sequence. The use of overlapping peptide fragments to determine the amyloidogenicity of individual residues ensures that the neighboring sequences of a given residue modulate its β-aggregation tendency. They have also run discrete molecular dynamics of the SOD1 dimer with square-well interaction potentials between beads (coarse-grained polypeptide model with four beads for the backbone atoms and two sidechain beads) according to contacts formed in the native state. Such native-structure-based Go model, derived using contact maps corresponding to the monomeric state and dimeric interface, was supplemented with an additional energy term to favor the formation of domain swapping interactions. A multiplicative coefficient in front of the domain swapping interactions was used for regulating the degree of intermonomer overlap, which represents the effective concentration of the protein. With these two very different polypeptide models and simulation protocols the authors identified the same amyloidogenic regions in the SOD1 sequence: the two termini, the β-strands 4 and 7, and the two crossover loops. A very interesting suggestion based on the simulation study of Dokholyan and coworkers [41•] is that a high aggregation propensity might result from the 'synergistic' placement of an amyloidogenic segment in a region of the protein that is most likely to form intermolecular contacts under conditions that destabilize the native state.

A simulation-based 'divide-and-conquer' approach was recently proposed by Cecchini and coworkers [16•] to obtain structural information on amyloidogenic polypeptides. The amino acid sequence was fragmented into overlapping segments, which were capped at each of the new termini. Decompositions into heptapeptide and hendecapeptide fragments were performed to investigate effects related to peptide length. Implicit solvent molecular dynamics simulations of oligomeric systems (three or six peptides in a box) were performed for each segment, starting from peptides well separated in space (i.e. without intermolecular contacts). The use of an implicit model of the solvent [42] allows for sampling of several association–dissociation events in a timescale of one microsecond, with each simulation requiring 10 or 25 days for three or six hendecapeptides, respectively. To validate the structural stability of the ordered aggregates observed in the implicit solvent runs, 50-ns control simulations with explicit water molecules and counterions were carried out for a subset of the segments. The combined implicit-solvent–explicit-solvent simulation approach provides information on the β-aggregation tendency along the polypeptide sequence (i.e. the β-aggregation profile) [16•]. Given the experimental evidence on in-register parallel arrangement of Aβ42 [4•] and the islet amyloid polypeptide [43], aggregation simulations of short stretches are a good approximation of the fibrillar environment and the observations made on the stretches can be extrapolated to full-length polypeptides. Thanks to the atomic detail information provided by the molecular dynamics simulations, the β-aggregation profile of Aβ42, the islet amyloid polypeptide and the N-terminal domain of the yeast prion-like protein Ure2p could be structurally characterized [16•]. For Aβ42, secondary structure analysis of the trajectories unveiled the presence of four turn-like sites: S8G9, G25S26, G29A30, and G38V39 (Figure 1). Interestingly, the location of the first three turns had been already suggested by solution NMR spectroscopy [44], solid-state NMR spectroscopy [45] and proline scanning mutagenesis [46], respectively. Although the four sites with turn propensity could have been detected by algorithms for secondary structure prediction, the consequences of such propensity within the context of an oligomeric system can be determined only by the atomistic simulation approach. The identified turn-like segments correspond to large drops in β-strand propensity and are located at the borders of aggregation-prone regions (see Figure 3 of [16•]). Hence, their specific position on the sequence determines the location and width of the aggregation hot-spots that are supposed to drive amyloid fibril formation, and to have an influence on

Figure 1



Structural model of part of the Aβ peptide obtained by molecular dynamics simulations of overlapping Aβ fragments [16•]. The most populated conformation of three H_{13}HQKLVFFAED$_{23}$ hendecapeptides is shown with carbon atoms in green, whereas the most populated conformation of three A_{21}EDVGSNKGAI$_{31}$ hendecapeptides is shown with carbon atoms in magenta. The relative orientation of the two fragments was obtained by overlapping the common part of the backbone (i.e. A_{21}ED$_{23}$). There is a striking similarity between this structural model and the one obtained by a combination of NMR spectroscopy and mutagenesis [4•].

the fibrillar conformation of Aβ42. After submission of the work by Cecchini and coworkers [16•], a 3D structural model of the Aβ42 fibril was obtained by using hydrogen-bonding constraints from quenched hydrogen–deuterium exchange NMR spectroscopy and sidechain packing constraints from pairwise mutagenesis studies [4•]. Notably, the 3D structural model shows parallel, in-register β-sheets formed by residues 18–26 and 31–42, and a loop at residues 26–30 in agreement with the molecular dynamics simulation results (compare Figure 1 with Figure 4c of [4•]).

Conclusions

Recent computational methods developed to improve the understanding of amyloid fibril formation include phenomenological models based on the physicochemical properties of the sidechains (e.g. β-propensity, hydrophobicity, aromatic content and charge), as well as atomistic simulation approaches. The former are simple, very efficient and can be applied to entire proteomes. Amyloid fibrils formed by a short stretch of a peptide or protein might have a different 3D structure than the fibril of the full-length sequence. Yet, the predictive ability of the models based on physicochemical properties [25•,26,27] and experimental evidence [2,15•,22] indicate that the amyloid-promoting part of a protein can be a short segment of the entire chain [47•]. These simple phenomenological models based on physicochemical properties are able to predict aggregation rates with reasonable accuracy and identify β-aggregation-prone fragments in proteins. The success of these simple models, which use only the protein sequence as input, is due to the regular structural arrangement and the important role of sidechains in β-sheet aggregates [14,15•,20,21,23].

Although they require large computational resources, the multiscale approaches based on molecular dynamics simulation techniques [16•,41•] provide structural information at atomic level and are, therefore, expected to have an important role in planning mutagenesis

experiments to modulate amyloid propensity. The ultimate goal of these simulation methods is to help in discovering small molecules which can interfere with formation and accumulation of the toxic species [48•].

Acknowledgements

I am very grateful to Marco Cecchini, François Marchand, Riccardo Pellarin and Gian Gaetano Tartaglia for very interesting and challenging discussions on amyloidogenic propensity, structure of ordered β-aggregates, and kinetics of fibril formation. They have implemented into software tools the models and methods developed in my research group. Most of the calculations needed for methodological development and applications were performed on the Matterhorn Beowulf cluster of the University of Zurich, and I thank F Dey, Ph Schuetz, C Bolliger and A Godknecht for computational help. This work was supported by grants from the Swiss National Science Foundation and the National Competence Center in Research in Neural Plasticity and Repair.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Sunde M, Blake C: **The structure of amyloid fibrils by electron microscopy and X-ray diffraction**. *Adv Protein Chem* 1997, **50**:123-159.

2. Nelson R, Sawaya MR, Balbirnie M, Madsen AO, Riekel C, Grothe R, Eisenberg D: **Structure of the cross-β spine of amyloid-like fibrils**. *Nature* 2005, **435**:773-778.

3. Ritter C, Maddelein ML, Siemer AB, Luhrs T, Ernst M, Meier BH, Saupe SJ, Riek R: **Correlation of structural elements and infectivity of the HET-sprion**. *Nature* 2005, **435**:844-848.

4. • Luhrs T, Ritter C, Adrian M, Riek-Loher D, Bohrmann B, Doeli H, Schubert D, Riek R: **3D structure of Alzheimer's amyloid-β(1–42) fibrils**. *Proc Natl Acad Sci USA* 2005, **102**:17342-17347.

The authors used quenched hydrogen–deuterium exchange NMR and mutagenesis studies to determine constraints on the fibrillar structure of the Alzheimer's Aβ(1–42) peptide. Structure calculation of a pentameric assembly using the experimentally determined restraints resulted in a β-strand (residues 18–26)–turn–β-strand (residues 31–42) motif. Interestingly, intermolecular sidechain contacts are formed between odd-numbered residues of the N-terminal strand and even-numbered residues of the C-terminal strand of the neighboring peptide.

5. • Pastor MT, Esteras-Chopo A, de la Paz ML: **Design of model systems for amyloid formation: lessons for prediction and inhibition**. *Curr Opin Struct Biol* 2005, **15**:57-63.

An interesting review article on the design and use of simplified systems (e.g. short peptides) to investigate amyloid-fibril formation. Both experimental and computational approaches are discussed.

6. Thirumalai D, Klimov DK, Dima RI: **Emerging ideas on the molecular basis of protein and peptide aggregation**. *Curr Opin Struct Biol* 2003, **13**:146-159.

7. Dokholyan NV: **Studies of folding and misfolding using simplified models**. *Curr Opin Struct Biol* 2006, **16**:79-85.

8. Gsponer J, Vendruscolo M: **Theoretical approaches to protein aggregation**. *Protein Pept Lett* 2006, **13**:287-293.

9. Chiti F, Stefani M, Taddei N, Ramponi G, Dobson CM: **Rationalization of the effects of mutations on peptide and protein aggregation rates**. *Nature* 2003, **424**:805-808.

10. 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.

11. de Groot NS, Aviles FX, Vendrell J, Ventura S: **Mutagenesis of the central hydrophobic cluster in Aβ42 Alzheimer's peptide. Side-chain properties correlate with aggregation propensities**. *FEBS J* 2006, **273**:658-668.

12. Dubay KF, Pawar AP, Chiti F, Zurdo J, Dobson CM, Vendruscolo M: **Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains**. *J Mol Biol* 2004, **341**:1317-1326.

13. Broome BM, Hecht MH: **Nature disfavors sequences of alternating polar and non-polar amino acids: implications for amyloidogenesis**. *J Mol Biol* 2000, **296**:961-968.

14. Lopez de la Paz M, Serrano L: **Sequence determinants of amyloid fibril formation**. *Proc Natl Acad Sci USA* 2004, **101**:87-92.

15. • Williams AD, Shivaprasad S, Wetzel R: **Alanine scanning mutagenesis of Aβ(1–40) amyloid fibril stability**. *J Mol Biol* 2006, **357**:1283-1294.

An alanine scanning mutagenesis of Alzheimer's Aβ(1–40) peptide was performed and for each mutant the free energy of fibril elongation was derived from the critical concentration value for fibril growth. Pro–Ala relative free energy values of fibril elongation were obtained by using previously measured Pro mutants. The results indicate that positions 18–21, 25–26 and 32–33 are particularly sensitive to the effects of Pro on the backbone structure. The lack of additivity in Ala double mutant studies suggests that the structure of Aβ fibrils is plastic.

16. • Cecchini M, Curcio R, Pappalardo M, Melki R, Caflisch A: **A molecular dynamics approach to the structural characterization of amyloid aggregation**. *J Mol Biol* 2006, **357**:1306-1321.

A molecular dynamics simulation protocol was developed to obtain structural information on ordered β-aggregation of amyloidogenic polypeptides (Figure 1). Notably, the β-aggregation propensity along the sequence of the Alzheimer's Aβ(1–42) peptide was found to be highly heterogeneous with a maximum at the segment V_{12}HHQKLVFFAE$_{22}$ and minima at four turn-like dipeptides. The predicted change in the aggregation propensity of a double-point mutant of the N-terminal domain of the yeast prion Ure2p was verified *in vitro* using the thioflavin T binding assay.

17. Tartaglia GG, 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.

18. Christopeit T, Hortschansky P, Schroeckh V, Guhrs K, Zandomeneghi G, Fandrich M: **Mutagenic analysis of the nucleation propensity of oxidized Alzheimer's β-amyloid peptide**. *Protein Sci* 2005, **14**:2125-2131.

19. Pellarin R, Caflisch A: **Interpreting the aggregation kinetics of amyloid peptides**. *J Mol Biol* 2006, **360**:882-892.

20. Azriel R, Gazit E: **Analysis of the structural and functional elements of the minimal active fragment of islet amyloid polypeptide (IAPP) An experimental support for the key role of the phenylalanine residue in amyloid formation**. *J Biol Chem* 2001, **276**:34156-34161.

21. Gazit E: **A possible role for pi-stacking in the self-assembly of amyloid fibrils**. *FASEB J* 2002, **16**:77-83.

22. Jones S, Manning J, Kad NM, Radford SE: **Amyloid-forming peptides from β(2)-microglobulin. Insights into the mechanism of fibril formation *in vitro***. *J Mol Biol* 2003, **325**:249-257.

23. Gsponer J, Haberthur U, Caflisch A: **The role of side-chain interactions in the early steps of aggregation: molecular dynamics simulations of an amyloid-forming peptide from the yeast prion Sup35**. *Proc Natl Acad Sci USA* 2003, **100**:5154-5159.

24. Cecchini M, Rao F, Seeber M, Caflisch A: **Replica exchange molecular dynamics simulations of amyloid peptide aggregation**. *J Chem Phys* 2004, **121**:10748-10756.

25. • Tartaglia GG, Cavalli A, Pellarin R, Caflisch A: **Prediction of aggregation rate and aggregation-prone segments in polypeptide sequences**. *Protein Sci* 2005, **14**:2723-2734.

A phenomenological model to predict aggregation rates and aggregation-prone segments was developed using results from molecular dynamics simulations of oligomeric systems of several heptapeptide sequences. The model, which did not require any parameter fitting, determined the parallel or anti-parallel β-sheet arrangement in the fibril and was shown to be able to identify the most amyloidogenic 5-residue segment in the 750-residue Alzheimer's amyloid protein precursor. Application to mammalian

and nonmammalian prion proteins provided insight into the species barrier.

26. Pawar AP, DuBay KF, Zurdo J, Chiti F, Vendruscolo M, Dobson CM: **Prediction of 'aggregation-prone' and 'aggregation-susceptible' regions in proteins associated with neurodegenerative diseases**. *J Mol Biol* 2005, **350**:379-392.

27. Fernandez-Escamilla AM, 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. Guerois R, Nielsen JE, Serrano L: **Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations**. *J Mol Biol* 2002, **320**:369-387.

29. • Linding R, Schymkowitz J, Rousseau F, Diella F, Serrano L: **A comparative study of the relationship between protein structure and β-aggregation in globular and intrinsically disordered proteins**. *J Mol Biol* 2004, **342**:345-353.

The authors applied TANGO to compare the β-aggregation propensity of globular proteins and a set of 296 intrinsically disordered proteins. The TANGO results indicate that the former contain almost three times as much β-aggregation prone regions as the latter. Similar β-aggregation propensity was observed in all-α, all-β and mixed α/β water-soluble proteins, as well as in membrane-associated proteins.

30. Rousseau F, Serrano L, Schymkowitz JWH: **How evolutionary pressure against protein aggregation shaped chaperone specificity**. *J Mol Biol* 2006, **355**:1037-1047.

31. Tartaglia GG, Pellarin R, Cavalli A, Caflisch A: **Organism complexity anti-correlates with proteomic β-aggregation propensity**. *Protein Sci* 2005, **14**:2735-2740.

32. Fowler SB, Poon S, Muff R, Chiti F, Dobson CM, Zurdo J: **Rational design of aggregation-resistant bioactive peptides: reengineering human calcitonin**. *Proc Natl Acad Sci USA* 2005, **102**:10105-10110.

33. Lopez de la Paz M, Goldie K, Zurdo J, Lacroix E, Dobson CM, Hoenger A, Serrano L: ***De novo* designed peptide-based amyloid fibrils**. *Proc Natl Acad Sci USA* 2002, **99**:16052-16057.

34. Lopez de la Paz ML, Lacroix E, Ramirez-Alvarado M, Serrano L: **Computer-aided design of β-sheet peptides**. *J Mol Biol* 2001, **312**:229-246.

35. Thompson MJ, Sievers SA, Karanicolas J, Ivanova MI, Baker D, Eisenberg D: **The 3D profile method for identifying fibril-forming segments of proteins**. *Proc Natl Acad Sci USA* 2006, **103**:4074-4078.

36. Kuhlman B, Baker D: **Native protein sequences are close to optimal for their structures**. *Proc Natl Acad Sci USA* 2000, **97**:10383-10388.

37. Sunde M, Serpell LC, Bartlam M, Fraser PE, Pepys MB, Blake CCF: **Common core structure of amyloid fibrils by synchrotron X-ray diffraction**. *J Mol Biol* 1997, **273**:729-739.

38. Gordon DJ, Balbach JJ, Tycko R, Meredith SC: **Increasing the amphiphilicity of an amyloidogenic peptide changes the β-sheet structure in the fibrils from antiparallel to parallel**. *Biophys J* 2004, **86**:428-434.

39. Yoon S, Welsh WJ: **Detecting hidden sequence propensity for amyloid fibril formation**. *Protein Sci* 2004, **13**:2149-2160.

40. Yoon S, Welsh WJ: **Rapid assessment of contact-dependent secondary structure propensity: relevance to amyloidogenic sequences**. *Proteins* 2005, **60**:110-117.

41. • Khare SD, Wilcox KC, Gong P, Dokholyan NV: **Sequence and structural determinants of Cu, Zn superoxide dismutase aggregation**. *Proteins* 2005, **61**:617-632.

To determine the amyloidogenic fragments in the enzyme Cu, Zn superoxide dismutase (SOD1) two different simulation techniques were used: discrete molecular dynamics with a coarse-grained Go model of a dimeric system and explicit water Newtonian dynamics of oligomeric systems. Notably, the two different simulation approaches identify identical regions of the SOD1 sequence as most amyloidogenic.

42. Ferrara P, Apostolakis J, Caflisch A: **Evaluation of a fast implicit solvent model for molecular dynamics simulations**. *Proteins* 2002, **46**:24-33.

43. Kajava AV, Aebi U, Steven AC: **The parallel superpleated β-structure as a model for amyloid fibrils of human amylin**. *J Mol Biol* 2005, **348**:247-252.

44. Hou L, Shao HY, Zhang YB, Li H, Menon NK, Neuhaus EB, Brewer JM, Byeon IJL, Ray DG, Vitek MP *et al.*: **Solution NMR studies of the Aβ(1-40) and Aβ(1-42) peptides establish that the met35 oxidation state affects the mechanism of amyloid formation**. *J Am Chem Soc* 2004, **126**:1992-2005.

45. Petkova AT, Ishii Y, Balbach JJ, Antzutkin ON, Leapman RD, 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 USA* 2002, **99**:16742-16747.

46. Williams AD, Portelius E, Kheterpal I, Guo JT, Cook KD, Xu Y, Wetzel R: **Mapping Aβ amyloid fibril secondary structure using scanning proline mutagenesis**. *J Mol Biol* 2004, **335**:833-842.

47. • Esteras-Chopo A, Serrano L, de la Paz ML: **The amyloid stretch hypothesis: recruiting proteins toward the dark side**. *Proc Natl Acad Sci USA* 2005, **102**:16672-16677.

The nonamyloidogenic protein α-spectrin Src homology 3 domain was converted into a β-aggregation prone sequence upon insertion of a six-residues amyloidogenic segment which did not perturb the stability of the folded state. Moreover, short stretches close to the inserted segment were also shown to belong to the protease-resistant core of the fibril. These experimental results provide strong evidence that the amyloidogenic part of a polypeptide sequence is located in short segments which can even promote the aggregation of adjacent stretches.

48. • Cohen T, Frydman-Marom A, Rechter M, Gazit E: **Inhibition of amyloid fibril formation and cytotoxicity by hydroxyindole derivatives**. *Biochemistry* 2006, **45**:4727-4735.

Three inhibitors of aggregation of the Alzheimer's Aβ(1–42) peptide were identified by screening 29 indole derivatives using fluorescence spectroscopy, atomic force microscopy and electron microscopy. Very interestingly, 4-hydroxyindole was shown to be the most effective inhibitor, whereas indole did not inhibit. This indicates that the understanding of the mechanism of inhibition requires investigations at the atomic level of detail.