Interactions in the native state of monellin, which play a protective role against aggregation†

Olga Szczepankiewicz,*[a] Celia Cabaleiro-Lago,[b] Gian Gaetano Tartaglia,[c] Michele Vendruscolo,[c] Thérèse Hunter,[d] Gary J. Hunter,[d] Hanna Nilsson,[a] Eva Thulin[b] and Sara Linse*[b]

Received 12th August 2010, Accepted 12th October 2010
DOI: 10.1039/c0mb00155d

A series of recent studies have provided initial evidence about the role of specific intra-molecular interactions in maintaining proteins in their soluble state and in protecting them from aggregation. Here we show that the amino acid sequence of the protein monellin contains two aggregation-prone regions that are prevented from initiating aggregation by multiple non-covalent interactions that favor their burial within the folded state of the protein. By investigating the behavior of single-chain monellin and a series of five of its mutational variants using a variety of biochemical, biophysical and computational techniques, we found that weakening of the non-covalent interaction that stabilizes the native state of the protein leads to an enhanced aggregation propensity. The lag time for fibrillation was found to correlate with the apparent midpoint of thermal denaturation for the series of mutational variants, thus showing that a reduced thermal stability is associated with an increased aggregation tendency. We rationalize these findings by showing that the increase in the aggregation propensity upon mutation can be predicted in a quantitative manner through the increase in the exposure to solvent of the amyloidogenic regions of the sequence caused by the destabilization of the native state. Our findings, which are further discussed in terms of the structure of monellin and the perturbation by the amino acid substitutions of the contact surface between the two subdomains that compose the folded state of monellin, provide a detailed description of the specific intra-molecular interactions that prevent aggregation by stabilizing the native state of a protein.

Introduction

The native state conformation of a protein is stabilized by the collective action of many non-covalent interactions including Coulomb and van der Waals interactions, hydrophobic effect and hydrogen bonds. These interactions, however, are also responsible for the formation of a variety of other states, stable or transient, that proteins can populate, including ordered aggregates, known as amyloid fibrils, which may be functional[1–3] or disease-associated.[4,5] Such amyloid fibrils are characterized by cross-β stacking of peptide segments perpendicular to the long axis of the fiber.[6,7] The ability to form amyloid fibrils may be a common characteristic of polypeptide chains, although the propensity to form such structures can vary dramatically depending on the intrinsic as well as extrinsic factors.[8] Intrinsic factors of the protein sequence that may favor or prevent aggregation include charge, hydrophobicity, spacing and patterns of polar and non-polar residues as well as secondary structure propensities.[9] There is evidence that protein sequences have evolved to avoid consecutive stretches of three or more hydrophobic residues,[10] or extended segments of alternating hydrophobic/hydrophilic residues.[11] A high positive or negative net charge hinders self-association,[12–14] while low α-helix propensity and high β-sheet propensity may promote amyloid formation.[9] In their natural environment, proteins have to avoid aggregation in a crowded environment that may contain up to 350 mg ml^{-1} of proteins as well as other biomolecules, ions, *etc.*[15–17] Extrinsic biological factors that cause or prevent aggregation include the interaction with cellular or extracellular components such as chaperones,[18–20] and the quality control mechanisms such as the ubiquitin–proteasome system.[21] Extrinsic physicochemical factors that affect aggregation and fibrillation properties are pH, temperature, ionic strength and protein concentration.[22–24]

The propensity of globular proteins to aggregate is often inversely related to the stability of the native state[4] and increased tendency for unfolding seems to promote fibrillation. Many polypeptide sequences contain local regions that are "sensitive" to aggregation. Single mutations in these regions can change the aggregation rate dramatically, while similar changes in other regions may have relatively little effect.[9,25] Protein aggregation is involved in many human disorders, including Alzheimer's, Creutzfeldt–Jacob, and Parkinson's diseases and dialysis-related amyloidosis. However, not every fibrous protein form is dangerous or disease causing. Collagen,

[a] *Lund University, Department of Biophysical Chemistry, Lund, Sweden. E-mail: Olga.Szczepankiewicz@bpc.lu.se*
[b] *Lund University, Dept. of Biochemistry, Lund, Sweden. E-mail: sara.linse@biochemistry.lu.se*
[c] *Dept. of Chemistry, University of Cambridge, Cambridge, UK*
[d] *Dept. of Physiology and Biochemistry, University of Malta, Malta*
† Electronic supplementary information (ESI) available: Text and seven Figures. See DOI: 10.1039/c0mb00155d

for example, is a fibrous protein that is the main structural component of bones, skin, ligaments and blood vessels in higher animals.[26] There are also many examples of functional amyloids in bacteria, yeast, plants and mammals; with functions ranging from biofilm formation, adhesion, scaffolding and information transfer.[1–3] For instance, the potato protein multicystatin (PMC) is a cysteine protease inhibitor that is stored in inactive aggregates and becomes active as an inhibitor when released in monomeric form upon reduction in pH or phosphate concentration. This protein has high structural homology with monellin,[27] the protein studied here.

In this work we investigated the interplay between the sequence and the structure in determining the aggregation process of single-chain monellin (scMN). Monellin was originally isolated from the tropical serendipity berry (*Dioscoreophyllum cumminsii*) and has been noted for its intense sweetness and tendency to aggregate.[28–33] Monellin is naturally composed of two polypeptide chains, MNA (45aa) and MNB (50aa), with high affinity for one another, and folds in a single globular domain with the β-grasp architecture.[34–36] In this work we considered a single-chain monellin (scMN) construct in which MNA is covalently linked to MNB. Folded monellin is about 100 000 times sweeter than sucrose on the molar basis. A single α-helix is packed perpendicularly on a five-stranded antiparallel β-sheet, and this is one of the most common folds known.[37,38] Recombinant scMN has retained sweetness and increased stability against thermal denaturation.[39] The structures of monellin and scMN superimpose, with MNA (or subdomain A) forming three of the β-strands and MNB (or subdomain B) forming two β-strands separated by the α-helix. The five scMN mutants studied here were retrieved from a phage display library based on their surface activity. All modifications are confined to subdomain A and alter the hydrophobic packing between the subdomains, and the surface charge of the protein. The scMN variant sequences were analysed by algorithms to identify the regions of the sequence that are likely to promote aggregation.[25] The response to heat treatment was studied using CD spectroscopy and thioflavin T (ThT) fluorescence was monitored to follow the fibrillation kinetics. We found a high aggregation tendency of the variants that was evident already at growth and purification and could only to low extent be abolished using co-expression with chaperones. A correlation is observed between the apparent midpoint of thermal denaturation and the lag phase for fibrillation, providing support to the view that there is a competition between folding and aggregation, which is ultimately caused by the fact that both the folded and the aggregated states are stabilized by similar interactions.[25]

Results

The stability and aggregation of the parent scMN and five of its mutational variants were investigated. Their amino acid sequences are presented in Fig. 1, and the gene sequences in expression plasmids are given in ESI† (Fig. S3). The mutations are introduced to alter the electrostatic charge of subdomain A or the hydrophobic interactions between subdomains A and B. The parent scMN contains many charged residues yielding a net charge of +2 at neutral pH (+3 on subdomain A and −1 on subdomain B) and extensive hydrophobic contacts in the interior between the two subdomains. Our parent scMN and its mutants contain the substitution C41S to avoid disulfide bonds. The mutants were retrieved from a phage display library of MNA based on their surface activity (see ESI†) and are extensively modified with a range of substitutions and truncations. All modifications are thus confined to subdomain A, and the molecular properties of the parent scMN and the mutants are presented in Table 1. The variants contain up to 14 substitutions and deletions (Fig. 1), which have changed the properties of the proteins dramatically compared to the parent scMN.

Expression and purification

Aggregation of all variants was apparent already at the stage of protein production. The mutant proteins were found in inclusion bodies while the parent scMN was present in the soluble fraction. We attempted co-expression of one mutant, scMN1, with the GroEL–GroES chaperonin system to minimize inclusion body formation. A time course experiment during which 1 ml aliquots were collected demonstrated that maximum expression was obtained as rapidly as 5 min after the addition of the inducer IPTG. SDS-PAGE analysis revealed that the large subunit of the chaperone GroEL co-aggregated with the mutant protein in the insoluble fraction of the cells. While most of monellin was observed in insoluble form together with GroEL, a small portion of the expressed monellin was observed as soluble protein in the supernatant (data not shown). This result is in agreement with a recent study which

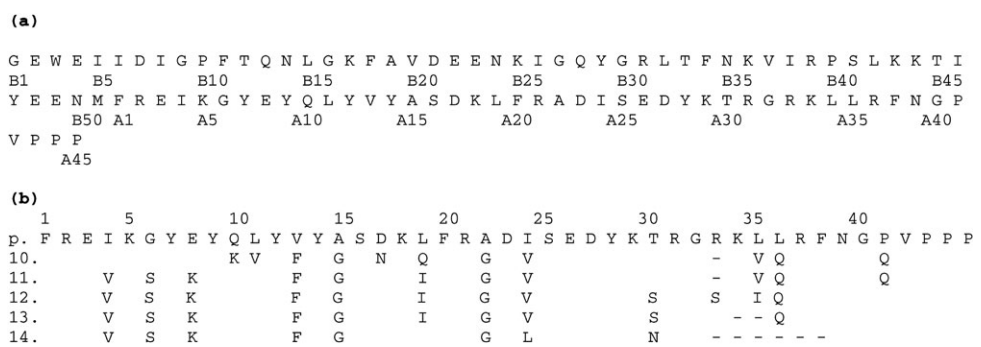

Fig. 1 (a) Amino acid sequence of parent scMN. (b) Amino acid sequences of the A fragment in the parent (p) and five mutated forms of scMN (numbered 10–14). Truncations are indicated by "-".

Table 1 Molecular properties of parent scMN and mutants

Protein	Net charge	Negative/positive[a]	Number of mutations	Number of truncations	Z^{agg}	$\tilde{Z}^{agg}$	Lag time/minutes	T_m/°C
Parent scMN	+2	14/16	0	0	0.63	0.19	>5600	72.2 ± 1.0
scMn10	+3	13/16	11	1	0.55	0.23	117 ± 5	47.4 ± 2.0
scMn11	+3	13/16	11	1	0.59	0.22	412 ± 19	52.9 ± 3.0
scMn12	+3	13/16	10	0	0.59	0.23	538 ± 21	51.5 ± 3.0
scMn13	+3	13/16	10	2	0.61	0.21	2078 ± 27	57.0 ± 2.0
scMn14	+1	13/14	8	6	0.58	0.20	2918 ± 100	62.3 ± 2.0

[a] Number of negative and positive residues in the sequence.

shows that unfolded monellin molecules bind very strongly to GroEL.[40]

Prediction of aggregation-prone regions

The intrinsic aggregation propensity scores, Z^{agg}, for the parent scMN and the mutants are summarized in Table 1; the larger the Z^{agg} score, the higher the aggregation propensity. The mutants have slightly lower Z^{agg} scores than the parent scMN.[25] The aggregation propensity scores with the structural correction, $\tilde{Z}^{agg}$, are also presented in Table 1. In this case, the mutants have higher $\tilde{Z}^{agg}$ scores than the wild type. These results indicate that the mutants, despite having a lower aggregation propensity than the wild type, are more prone to aggregation under native conditions since their amyloidogenic regions are less protected by the folded structure. In Fig. 2 the intrinsic aggregation propensity profiles, *i.e.* the sequence-dependent Z_i^{agg} scores[25] and the aggregation propensity profiles with the structural corrections, *i.e.* the sequence-dependent, $\tilde{Z}_i^{agg}$ scores are presented for the parent scMN and the mutants. Two regions are characterized by Z_i^{agg} values greater than 1 and are likely to be involved in the aggregation process: B4-6, which are Glu-Ile-Ile in both the parent and mutants, and A8-13, which has variable sequence over the mutants (Fig. 1–3). These two regions are not close in space in the folded structure of monellin. The aggregation-prone region in the A-fragment is situated in the first β-strand of the A fragment, while the aggregation-prone region in the B-fragment is placed in the first β-strand of the B fragment. The aggregation propensity profiles with structural corrections, $\tilde{Z}_i^{agg}$, are all lower than the profiles of intrinsic aggregation propensities, Z_i^{agg} (Fig. 3). This result is expected, since $\tilde{Z}_i^{agg}$ describes the

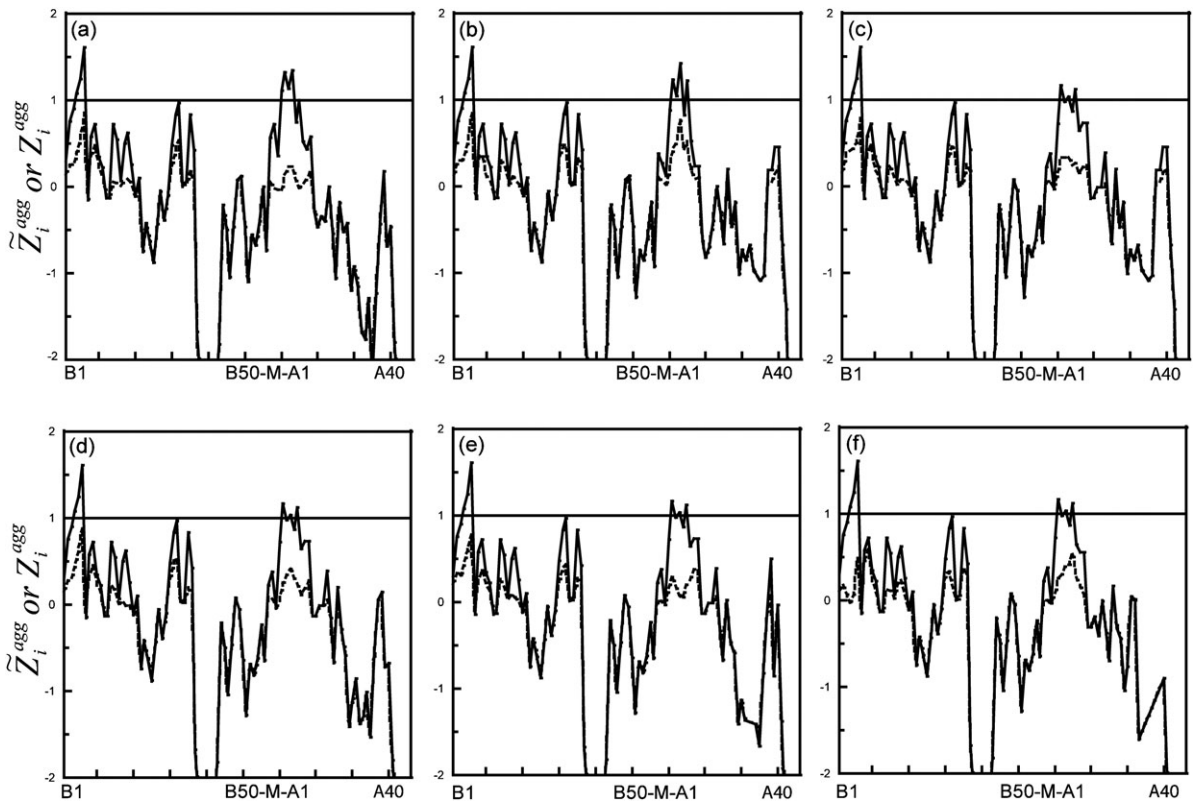

Fig. 2 The intrinsic, Z_i^{agg} (solid line) aggregation propensity profiles and the aggregation propensity profiles with structural modification, $\tilde{Z}_i^{agg}$ (dashed line), for (a) parent scMN, (b) scMN10, (c) scMN11, (d) scMN12, (e) scMN13 and (f) scMN14.

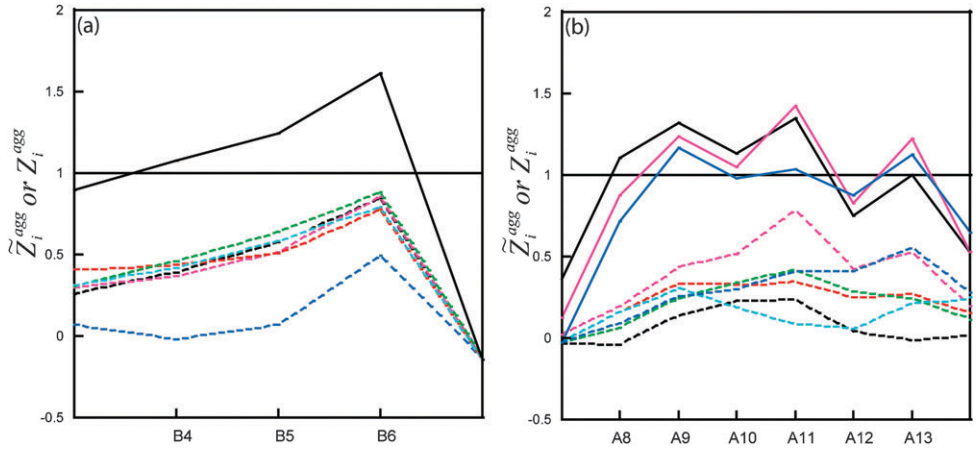

Fig. 3 Intrinsic aggregation propensity profiles, Z_i^{agg} (solid line), and aggregation propensity profiles with structural modification, $\tilde{Z}_i^{agg}$ (dashed line) for regions in scMN chain B (a) and chain A (b) where Z_i^{agg} is above 1. $\tilde{Z}_i^{agg}$ is shown for parent scMN (black), scMN10 (pink), scMN11 (red), scMN12 (green), scMN13 (cyan) and scMN14 (blue). All variants have the same sequence in region B4-6, and scMN variants 10, 11, 13 and 14 have the same sequence in region A8-13, thus only Z_i^{agg} of parent (black) is shown in (a) and only for parent, scMN10 (pink) and scMN12 (blue) in (b).

reduction in aggregation propensity caused by the burial of the aggregation-prone regions within the native structure. The results presented in Table 1 and Fig. 2 and 3 indicate that such protection is less effective in the mutants.

CD spectra

The structure and stability of the produced single-chain mutants were investigated using CD-spectroscopy. The spectra of the parent scMN and all mutants are shown in Fig. 4a. The concentrations are based on absorbance. The parent scMN displays a minimum around 215 nm of *ca.* −8600 deg cm^2 decimole^{-1}, which corresponds well to the signal for a mainly β-sheet protein. For all mutants the spectra reflect a folded protein. We note that the presence of a helical segment in the otherwise β-sheet dominated protein is more apparent in the spectra of mutants (characteristic double dip at 208 and 222 nm), than in the parent spectrum. The differences in the amino acid sequences could also affect the appearance in the spectra.

The near UV spectra in Fig. 4b show that the mutants are folded but have less tertiary structure compare to the parent scMN. The existence of the non-zero signal indicates that the mutants are folded with confined aromatic side chains. The mutation V13F in all of the mutants contributes to the near UV signal and therefore, no direct comparison with the parent sequence can be made.

Thermal denaturation by CD spectroscopy

We followed the temperature denaturation process by recording the CD signal at 208 nm from 4 to 96 °C for all mutants in comparison with the parent scMN (Fig. 5). For the parent scMN and all mutants we observe a transition towards lower signal intensity and the transition occurs at lower temperature for all mutants compared to the parent. The thermal denaturation of scMN was previously shown to be reversible when the protein was briefly heated to high temperature.[41,42] In earlier studies, we have observed that the reversibility of the parent scMN depends on protein concentration and prolonged heating

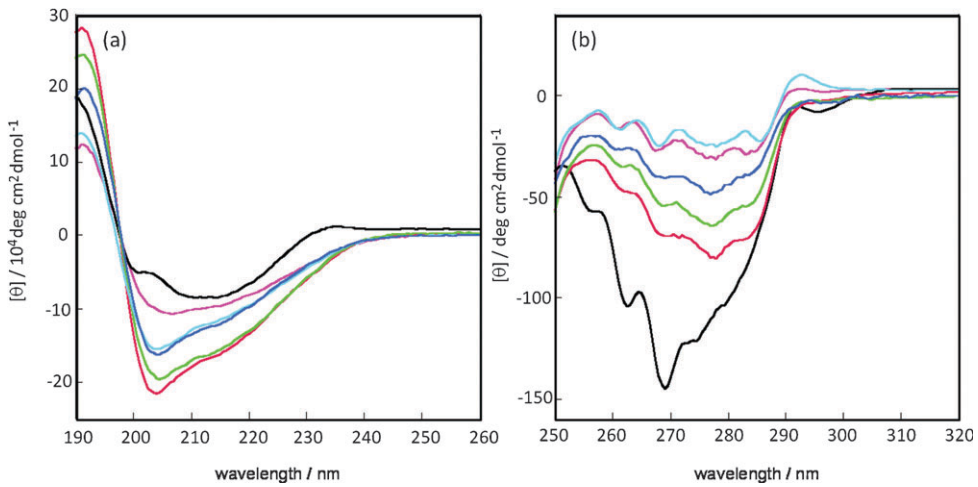

Fig. 4 Far UV (a) and near UV (b) CD spectra at 4℃ for parent scMN (black), scMN10 (pink), scMN11 (red), scMN12 (green), scMN13 (cyan) and scMN14 (blue).

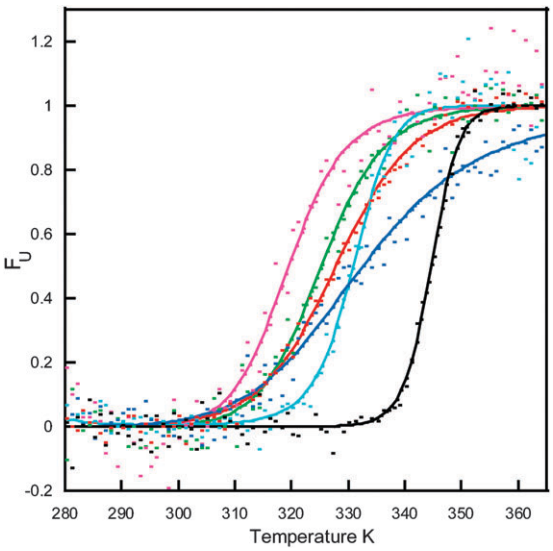

Fig. 5 Thermal stability as monitored by the CD signal at 208 nm. The data are normalized to show the unfolded fraction as a function of temperature for parent scMN (black), scMN10 (pink), scMN11 (red), scMN12 (green), scMN13 (cyan) and scMN14 (blue).

at high temperature leads to aggregation and precipitation.[43] In our experiments the scans were run up to 96 °C to maximize the post-transition base line. However, this reduced the reversibility of the denaturation. When the backward scans were run from 96 to 4 °C, the signal of parent monellin was fully regained, but not at all or only partially for the mutants (data not shown). Spectra recorded after cooling to 4 °C after denaturation show that denaturation is reversible for the parent but not for the mutants (data not shown). Thus fitting of eqn (6) to the data yields T_m for the parent scMN but only extract apparent T_m values for the mutants. For the parent scMN we obtain T_m of 72.2 ± 1.0 °C in agreement with earlier reports.[39,43] The mutants display transitions at apparent T_m values of; 47.4 ± 2.0 °C for scMN10, 52.9 ± 3.0 °C for scMN11, 51.5 ± 3.0 °C for scMN12, 57.0 ± 2.0 °C for scMN13 and 62.3 ± 2.0 °C for scMN14.

Fibrillation kinetics

The fibrillation of single-chain monellin was followed by monitoring ThT fluorescence, which increases markedly upon binding of ThT to large aggregates or fibrils (Fig. 6). The parent scMN does not appear to fibrillate during the time course of our experiment (up to 96 hours) under the chosen conditions (10 mM sodium phosphate buffer, 0.02% NaN_3, pH 3.5). Mutant scMN10 has a lag phase of 117 ± 5 minutes. scMn11 and scMn12 have a lag phase in the range of 400–600 minutes. Mutants scMN13 and scMN14 fibrillate with a much longer lag phase of 2000 minutes or more (Fig. 6). All the fibrillating monellin mutants display a lag phase followed by a rapid elongation phase.

TEM

TEM images after completion of the fibrillation experiments reveal that fibrils are formed for scMN10-14, while no fibrils are observed for the parent scMN (Fig. 7).

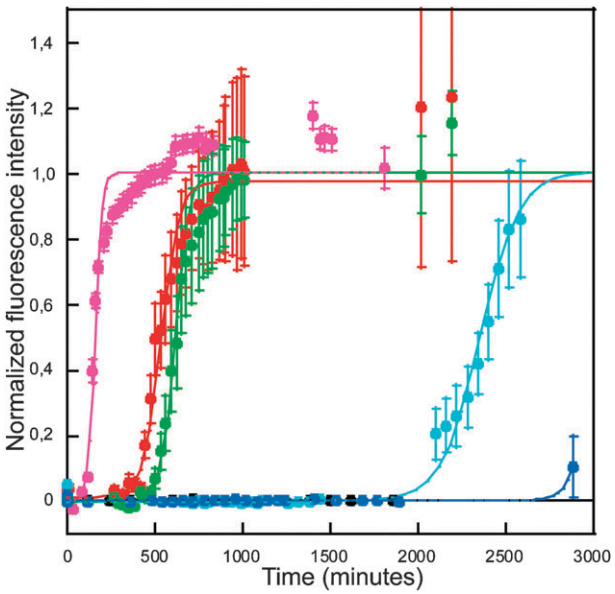

Fig. 6 Fibrillation kinetics of scMN parent and mutants (20 μM) at 37 °C and pH 3.5, monitored by temporal development of ThT fluorescence. The normalized fluorescence intensity is shown for parent scMN (black), scMN10 (pink), scMN11 (red), scMN12 (green), scMN13 (cyan) and scMN14 (blue) as averages and standard deviations of 6 measurements.

Discussion

The results of the present work reveal that there is an increased risk of amyloid formation of single chain monellin when the interactions within the folded state of the protein are perturbed. An important factor responsible for this increased tendency for fibril formation thus appears to be a destabilization of the protein leading to increased exposure of aggregation-prone segments. Reduced stability towards thermal denaturation is indeed found for the mutants; the onset of denaturation occurs at lower temperature and the apparent mid-point of denaturation is reduced in all mutants relative to the parent protein. A correlation (R = 0.96) is found between protein stability (Fig. 5) and aggregation (Fig. 6), more precisely between the apparent mid-point of the thermal denaturation and the lag time for fibrillation (Fig. 8a). Clearly, the least stable mutants display the shortest lag time for fibrillation, mutants with intermediate stability have intermediate lag times, and the most stable mutant has the longest lag time. The parent scMN, which is more stable than all the mutants, has a lag time that is longer than the time frame of the study (6000 minutes). These results are consistent with previous studies that also reported the presence of a correlation between protein stability and fibrillation time for a series of mutants of immunoglobulin light chain protein (R = 0.88)[44] and acylphosphatase (R = 0.93).[45] In order to further support these conclusions we calculated the aggregation propensities using a method that takes account of the protective role of the native state of a protein against aggregation.[25,46] The method is based on the observation that aggregation-prone regions are often unable to initiate the aggregation process because they are buried within the folded structure of a protein. By considering this effect we calculated the aggregation propensity scores with

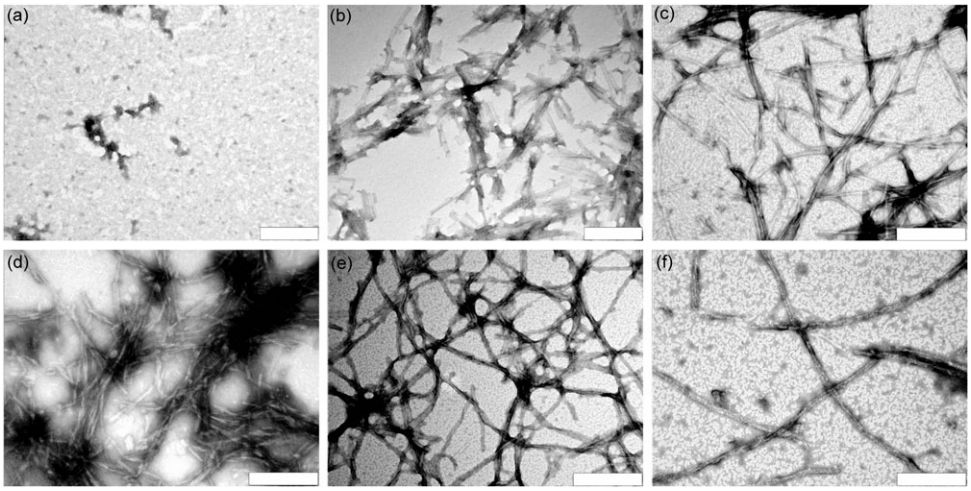

Fig. 7 Negatively stained Transmission Electron Microscopy images of scMN parent and mutants. Images were taken at the end of the fibrillation experiment for (a) parent scMN, (b) scMN10, (c) scMN11, (d) scMN12, (e) scMN13, (f) scMN14. Scale bar indicates 200 nm.

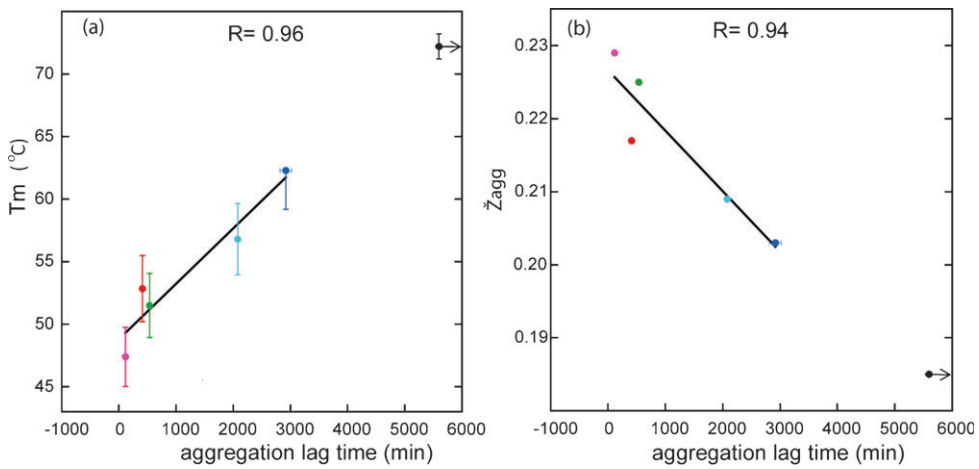

Fig. 8 (a) Correlation between apparent midpoint of thermal denaturation, T_m, and aggregation lag time for parent scMN and mutants. (b) Correlation between predicted aggregation propensity with structural correction, $\check{Z}^{agg}$, and aggregation lag time and of parent scMN and mutants. Parent scMN (black), scMN10 (pink), scMN11 (red), scMN12 (green), scMN13 (cyan) and scMN14 (blue).

the structure correction, of the parent scMN and of the five mutational variants considered in this study, finding an overall high correlation with the aggregation lag time (Fig. 8b).

Taken together, our results indicate that the amino acid sequence of monellin contains aggregation-prone segments that are protected from exposure in the wild-type by stabilizing interactions that strongly favor the folded state of the protein over unfolded or partially unfolded states.

In order to better understand which interactions in the parent protein that are perturbed by the mutations, we analyzed the crystal structure of wild type monellin. In this structure, the A and B chains are packed closely and interact by both intermolecular hydrogen bonds and hydrophobic interactions.[37] Positions that are altered or are missing in all mutants include V13, A15, A22, I24, L35 and L36 (Fig. 1). V13, A15, A22, I24, and L35 are in close contact (within 5 Å) with the B chain and seem to have important interactions with the helix and one strand of the B chain (Fig. 9 and 10, Table 2). Mutation or deletion of these amino acids may therefore decrease the affinity between the two chains and result in lower stability towards unfolding. L36, by contrast, is situated on the surface of the monellin (Fig. 9 and 10). Even though L36 does not seem to have any close contact with the B chain, it is involved in intra-domain interactions within the A chain that may be of great importance for protein stability (Fig. 11 and Table 2). This is true also for residues V13, A15, A22, I24, and L35 (Fig. 11 and Table 2). This analysis points to multiple non-covalent interactions involving V13, A15, A22, I24, L35 and L36 that may favor the folded over unfolded protein. Weakening of these non-covalent interactions may lead to destabilization of the protein, increased exposure of an aggregation-prone segment and thereby enhanced aggregation and formation of amyloid fibrils.

The variations in stability and fibrillation rate among the mutants (all of which contain substitutions or deletion of V13, A15, A22, I24, L35 and L36) are likely to result from a complex combination of all non-covalent interactions within and between subdomains, β-sheet propensity, and the

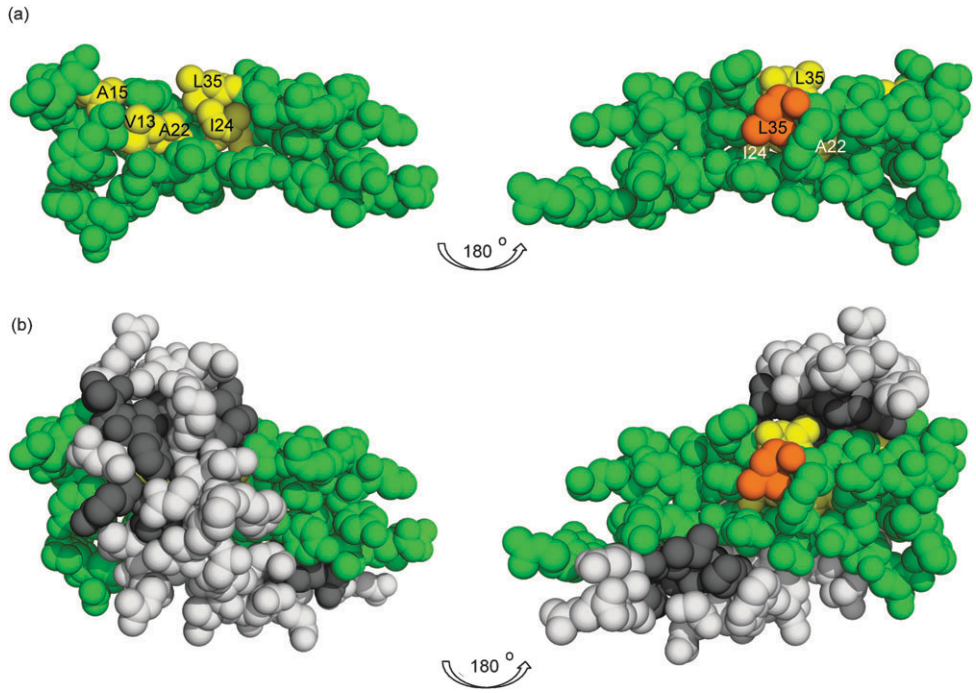

Fig. 9 Positions in chain A that are altered or deleted in all mutants (yellow or orange). In yellow, amino acids in close contact (< 5 Å) with chain B (V13, A15, A22, I24, L35) and in orange, L36 which is not in contact with chain B. (a) Chain A in two orientations related by a 180° rotation. (b) MN with the B chain in grey in two orientations related by a 180° rotation. The amino acids in chain B that are in close contact with V13, A15, A22, I24 and/or L35 are colored in dark gray. This figure was prepared using the structure 4MON.pdb and the program PyMOL (DeLano Scientific).

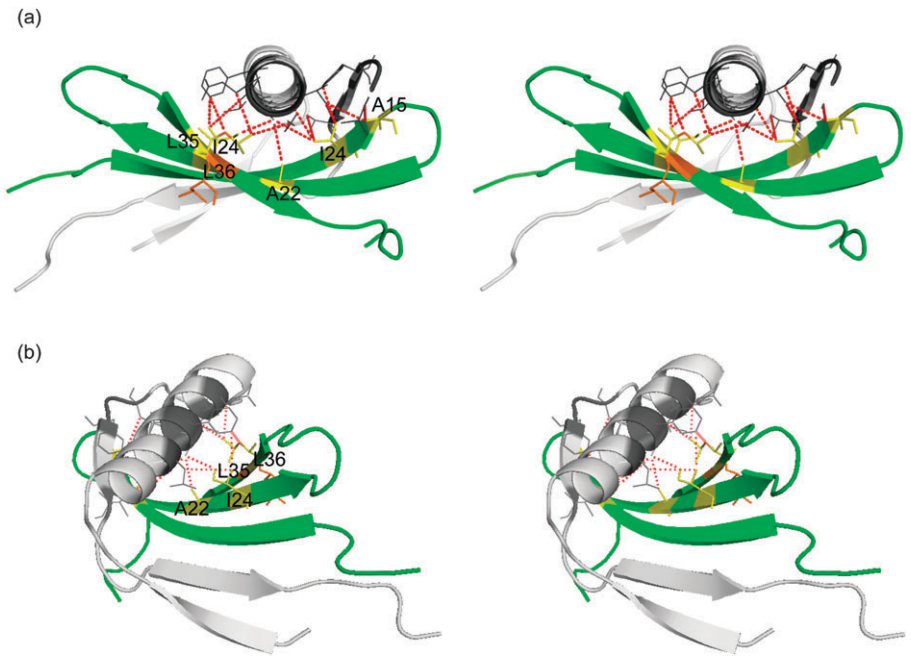

Fig. 10 Stereo images showing close contacts (< 5 Å) between side chains in chain B (grey ribbon) and residues V13, A15, A22, I24 and/or L35 in chain A (green ribbon). Color coding of amino acids agrees with Fig. 9. (a) and (b) show two different orientations.

aggregation propensity of the sequences. One factor that may modulate the general effects due to the above substitutions may be the variation in net charge and charge distribution over the mutants. High net charge of a protein may prevent aggregation due to inter-protein repulsion; at the same time, a high net charge may interfere with protein folding and reduce the stability of the folded relative to unfolded states and increase the exposure of aggregation-prone regions.[14] These two effects work in opposite direction and prevent and promote aggregation, respectively. Among the mutants, the

Table 2 Summary of the inter- and intra-chain interactions of the amino acids in fragment A

Pos in A	Residues in chain B (within 5 Å)	Intra-chain interactions A–A (within 5 Å)
E3	—	R3, I4
I4	T45(β3), I46, Y47	Y7(β4), Y28(L5-6)
E8(β1)	K44(β1), I46	D23(β5), S25(β5), L36(β6)
Q10(β1)	W3(β1), M42(β3)	R21(β5), D23(β5)
L11(β1)	T12(H), L15-G16(H), V37(β2), C41(β3)	A22(β5), I24(β5),
V13(β 1)	G16(H), A19(H), V20(H), F34(β2), K36-V37(β2)	F20(β5), **A22**(β5), F38(β6)
A15(β1)	V20(H), L32-T33(LH-β2), F34(β2)	F20(β5), F38(β6)
D17(L1-2)	N35(β2)	
L19(β2)	I38(LH-β2)	Y12(β4), Y14(β4), V42(β6), P44(β6), P45(β6)
A22(β2)	A19(H)	L11(β4), **V13**(β4), **L35**(β6), F38(β6)
I24(β2)	L15(H), F18-A19(H)	Y9(β4), L11(β4), R33(β6), **L35**(β6)
D27(β2)	—	T30(L5-6), R31(L5-6), G32(L5-6)
T30(L2-3)	—	D27(β5)
R33(β3)	L15(H), F18(H)	E26(β5)
L35(β3)	F18-A19(H), E22-E23(H)	**A22**(β5), **I24**(β5)
L36(β3)	—	E8(β4), D23(β5), S25(β5)

βX corresponds to a specific β strand. Strands 1–3 are in chain B and 4–6 in chain A. LX-Y corresponds to a loop between two secondary structures. H corresponds to the helix in chain B. In bold are the positions that are altered in all mutants.

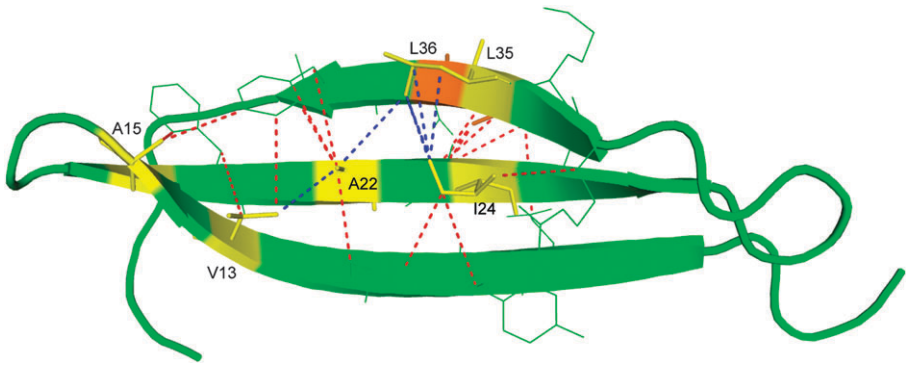

Fig. 11 Intereo images showing close contacts (<5 Å) between side chains in chain B (grey ribbon) and residues V13, A15, A22, I24 and/or L35 in chain A (green ribbon). Color coding of amino acids agrees with Fig. 9. (a) and (b) show two different orientations.

most stable and slowest fibrillating one (scMN14) has a net charge of +1, while the intermediate and fast ones (scMN10-13) have a net charge of +3. In the case of scMN14, the dominating effect of charge seems to be increased aggregation due to reduced electrostatic repulsion between protein molecules. In the other mutants, the dominating effect of charge on aggregation seems to be a destabilization of the folded relative to unfolded state and increased exposure of the amyloidogenic segment. These results suggest that the parent scMN has the most optimal charge balance to avoid aggregation. The β-sheet propensity may also affect fibril formation. The two fastest fibrillating mutants scMN10 and scMN11 have one more β-branched amino acid than the parent (V13F, L35V and L11V or L19I), while the two slowest fibrillating mutants have one or three less β-branched residues than the parent (scMN13: V13F, L19I, T30S; scMN14: V13F I25L, and T30N), and the intermediate mutant has zero net change (scMN12: V13F, L19I, L35I, and T30S). The role of sequence truncations is more difficult to evaluate, but we note that the longest truncation of six residues (R33-F38) is found in the slowest mutant (scMN14), and a two-residue truncation (K34-L35) in the next slowest (scMN13).

Structural studies of amyloid fibrils have primarily focused on short peptides (Table 3). Many of these amyloid-forming

Table 3 Short peptides shown to form amyloid fibrils

Sequence of peptide	Origin	Ref.
KLVFFAE	Amyloid-β protein	53
NFGAIL	Islet amyloid protein	54
KFFEAAAKKFFE	Design	55
AAXK	α-Synuclein	56
DWSFYLLYYEFT	β-2 Microglobulin	57
SFFSFLGEAFD	Serum amyloid protein	58
FAIRHF and FENKFAV	Abri	59
GNNQQNY	Yeast protein Sup35	60

peptides have aromatic and charged residues thought to play an important role in fibrillogenesis,[47] although high prevalence of hydrophilic residues is also compatible with amyloid formation.[7] The computational analysis of the aggregation propensity of monellin identified residues 8–13 of the A-chain (EYQLYV in parent ScMN) as a potentially aggregation prone region (Fig. 2 and 3). The combination of hydrophobic and aromatic side chains and one negative charge is the most striking feature of this sequence, which shows resemblance with several of the reported amyloidogenic peptides (Table 3). Also B4-6 (EII in both the parent and mutants) has aggregation score above 1, and is located close to the N-terminus of scMN. We cannot from the present data deduce the specific

role of these two regions (B4-6 or A8-15) in the fibrillation process of the mutants. Amyloid fibrils have previously been observed for MNB but not for MNA.[30] The α-helix in free MNB, or in MNB in native two-chain monellin seemed to be converted to a β-sheet conformation in the amyloid formation process, however, MNB lost its ability to form amyloid when covalently connected to the A-chain in scMN[30] presumably due to the structural protection as discussed above.

Experimental methods

Chemicals

All chemicals were of analytical grade.

Gene cloning and protein expression

Pet 3a vectors containing synthetic genes coding for the scMN variants were prepared by polymerase chain reaction (PCR), cutting with restriction enzymes and cloning using standard procedures. The gene for the parent scMN codes for a Met residue followed by MNB (with substitutions C42S and M42L), another Met residue and MNA.[43] The mutation BC41S was introduced to avoid S–S bridge formation and the mutation BM42L was introduced to exclude Met in the B sequence since the linking Met residue between B and A chains allows for CNBr cleavage to produce the A and B chains (Fig. 2). The Pet 3a vectors, with the gene for parent or variant scMN, were transformed into *E. coli* of strain BL21 DES3 PlysS star (Invitrogen, Groningen, The Netherlands) by electroporation and spread on LB/agar plates with 50 μg ml^{-1} ampicillin and 30 μg ml^{-1} chloramphenicol. Single colonies were picked and inoculated overnight in a LB medium containing 50 μg ml^{-1} ampicillin and 30 μg ml^{-1} chloramphenicol. The overnight cultures were used to inoculate, with a ratio of 1 : 100, 500 ml LB cultures containing the same antibiotics. The protein production was started by addition of 0.4 mM IPTG after OD_{600} of the cultures reached 0.6. The cells were harvested by centrifugation at 6000 *g* for 5 min, 3–4 hours after induction, resuspended in 10 mM MES buffer, pH 6.0 and stored at −80 °C.

Purification of parent scMN

The cell suspensions were thawed and sonicated for 5 min and subsequently centrifuged at 12 500 *g* for 10 min. The parent scMN was in the supernatant and was purified by ion exchange chromatography on a CM-cellulose column pre-equilibrated in 10 mM MES buffer, pH 6.0. The protein was eluted with a linear salt gradient from 0–0.5 M NaCl, and purity confirmed by native agarose gel electrophoresis and SDS-PAGE.

Purification of scMN variants

The cell suspensions were thawed and sonicated for 5 min and subsequently centrifuged at 12 500 *g* for 10 min. The scMN variants were in the pellet, which was dissolved in 8 M urea, 10 mM ammonium acetate, pH 4.8, stirred for 1 h and centrifuged at 12 500 *g* for 10 min. The resulting supernatant was subjected to one or two ion-exchange steps. First it was loaded onto a CM-cellulose column pre-equilibrated in 10 mM ammonium acetate buffer, pH 4.8. The column was washed with a salt gradient 0–1 M NaCl followed by 8 M urea in 10 mM ammonium acetate, pH 4.8, and the proteins were eluted with a linear salt gradient from 0–0.5 M NaCl in 8 M urea, 10 mM ammonium acetate, pH 4.8. The absorbance at 280 nm was used to identify the protein containing fractions and these were pooled and dialyzed against ddH_2O for two days. The pH of the dialysate was adjusted to 3.5 and the sample was analysed by native agarose gel electrophoresis and SDS-PAGE and freeze-dried. The variant scMN14 was judged to be sufficiently pure after this step (more than 90% pure on protein level), and the amount was 8.9 mg. The other variants were dissolved at 1 mg ml^{-1} in 8 M urea and pumped through a DEAE cellulose column pre-equilibrated in 8 M urea, 10 mM Tris/HCl, pH 7.5 to remove anionic impurities. The same buffer with 8 M urea was pumped onto the column until all scMN variants were eluted as judged by absorbance at 280 nm. The protein containing fractions were pooled and dialyzed against ddH_2O for two days. The pH of the dialysate was adjusted to 3.5 and the samples were freeze-dried. The purity of the scMN variants was confirmed by native agarose gel electrophoresis and SDS-PAGE. Around 40–140 mg of pure protein was obtained for the variants scMN10–13 from 2 l of culture. The purity after the last purification step was judged to be more than 90% on the protein level.

Secondary structure and thermal denaturation by CD spectroscopy

CD spectroscopy was performed using a JASCO J-720 and JASCO J-815 spectropolarimeter with a JASCO PTC-343 and a JASCO PTC-4235/15 Peltier-type thermostated cell holder. The secondary structures of the parent scMN and the variants were investigated by recording far-UV CD spectra between 190 and 250 nm at 4 °C for 1 μM protein in 10 mM phosphate buffer, pH 3.5. This pH was chosen because of poor solubility of the mutants at higher pH. Thermal unfolding was monitored by recording the CD signal at 208 nm from 4 °C to 96 °C for 5.0 μM protein in 10 mM phosphate buffer, pH 3.5. The scan rate was 1 °C min^{-1} and the signal was recorded every 1 °C using a response time of 16 s. Quartz cuvette with 10 mm path length was used. Buffers were filtered through 0.2 μm filters to remove possible particulate impurities, and the protein samples were centrifuged for 10 min at 13 000 rpm to remove aggregates.

The stability of the parent and mutant scMN was modeled using a two-state model with the equilibrium constant, *K*, between the native and unfolded forms:

$$N \leftrightarrow U \quad K = \frac{[U]}{[N]} \tag{1}$$

The fraction unfolded, F_U, is:

$$F_U = \frac{K}{1+K} = 1 - F_N. \tag{2}$$

where F_N is the native fraction.

The free energy of unfolding is then:

$$\Delta G^{o}_{NU} = -RT \ln K \tag{3}$$

The temperature dependence of ΔG°_{NU} is:

$$\Delta G^{\circ}_{NU}(T) = \Delta H^{\circ}_{NU}(T_m)\left[1 - \frac{T}{T_m}\right] + \Delta C_p\left[T - T_m - T\ln\frac{T}{T_m}\right] \quad (4)$$

where R is the gas constant and T_m is the temperature at midpoint of transition. For temperature denaturation measured by CD, the fraction of unfolded protein is:

$$F_U = \frac{Y - Y_N}{Y_U - Y_N} \quad (5)$$

Using eqn (2), (3) and (5), assuming linear pre- and post-transition baselines Y_N and Y_U, the CD signal Y is:

$$Y(T) = \frac{Y_N(T) + Y_U(T)\exp\left[\frac{-\Delta G^{\circ}_{NU}(T)}{RT}\right]}{1 + \exp\left[\frac{-\Delta G^{\circ}_{NU}(T)}{RT}\right]} \quad (6)$$

This eqn (6) is fitted to the CD denaturation data.

Co-expression of the mutant scMN10 with chaperones

1 μl of the construct of scMN10 in the Pet 3a (Amp^R) vector (around 1 μg μl^{-1}) was mixed together with 4 μl of pTETGroESL vector (5 μg μl^{-1}). pTETGroESL is a tetracycline derivative of the pREP4-GroESL vector encoding the GroEL and GroES operon obtained from Dr Steiger (Hoffman La Roche Ltd).[48] Calcium competent BL21 DE3 (from Novagen) cells were prepared according to standard procedures and 100 μl of the cells were transformed with this DNA mixture.[49] Cells were kept on ice for 30 min then heat shocked for 3 min at 42 °C. The cells were then mixed with 400 μl of 2TY media (50 ml, 16 g l^{-1} tryptone, 10 g l^{-1} yeast extract and 5 g l^{-1} NaCl), and incubated for one hour at 37 °C. The cell suspension was spread on agar plates supplemented with 25 μg ml^{-1} tetracycline and 25 μg ml^{-1} ampicillin to obtain well separated colonies of transformed cells. Various ratios and dilutions of the two co-transformed vectors were used to obtain the best results. The plates were incubated at 37 °C overnight and then investigated. The presence of both antibiotics, tetracycline and ampicillin ensured that the resulting transformants harboured both the Pet 3a vector with the mutant construct and the pTet vector containing the GroESL construct. A selected number of colonies were chosen and cultured in 2TY media also supplemented with 25 μg ml^{-1} tetracycline and 25 μg ml^{-1} ampicillin at 37 °C. When the OD_{600} reached 0.4, 0.1 mM IPTG was added and the cultures were grown overnight at 30 °C. Cell samples were taken out at different time points and centrifuged for 15 min at 3000 rpm at 4 °C. The cells were then lysed with B-Per^R bacterial Protein Extraction reagent (Pierce) according to the manual of this reagent. The lysate was centrifuged for 30 min at 13 000 rpm to remove the cell debris from the supernatant. Samples of the cell debris (insoluble fraction) and supernatant (soluble fraction) were analysed on 15% SDS-PAGE gels.

ThT fluorescence assay

Aggregation kinetics were monitored using thioflavin T (ThT; Calbiochem) as a reporter. Fluorescence signal was measured at regular intervals using a Molecular Devices SpectraMax M2 microplate reader (Sunnyvale, CA, USA) with excitation and emission at 440 nm and 480 nm, respectively. Each experimental point is an average of the fluorescence signal of 6 wells containing aliquots of the same solution. Samples for aggregation experiments were prepared as follows. Lyophylised protein was dissolved in 10 mM sodium phosphate buffer, 0.02% NaN_3, pH 3.5 to a concentration of 1 mg ml^{-1}. Solution was vortexed and centrifuged to eliminate any solids. After dilution in the same buffer, the solution was aliquoted in a 96 well black fluorescence plate, NUNC 96 black Polypropylene MicroWell™ Plates, and ThT stock solution was added. Typical concentrations in these aggregation experiments are 20 μM protein and 18 μM ThT. Plates were incubated at 37 °C and shaken at 700 rpm using a VorTemp 56™ Incubator/Shaker with an orbit of 3 mm (Labnet International, Berkshire, UK). Kinetic parameters were obtained by fitting the following experimental sigmoidal equation to the average experimental curves.[50,51]

$$Y = y_0 + \frac{y_{max} - y_0}{1 + e^{-(t - t_{1/2})k}} \quad (7)$$

In eqn (7), Y is the fluorescence intensity at time t, y_0 and y_{max} are the initial and maximum fluorescence intensities, respectively, $t_{1/2}$ is the time required to reach half the maximum intensity and k is the apparent first-order aggregation constant. Lag time is defined as

$$\text{lagtime} = t_{1/2} - 2/k \quad (8)$$

Transmission electronic microscopy

Negative staining TEM samples were prepared as follows: 10 μL of sample was applied to a formwar/carbon coated grid for 2 min and blotted with a filter paper. The sample was stained with an aqueous solution 2% uranyl acetate for 2 min, blotted and air dried. Images were obtained using a JEOL 2000 Electron Transmission Microscope operated at 80 V.

Prediction of aggregation propensities

The intrinsic aggregation propensity, p_i^{agg}, of an individual amino acid is defined as[25]

$$p_i^{agg} = \alpha_h p_i^h + \alpha_s p_i^s + \alpha_{hyd} p_i^{hyd} + \alpha_c p_i^c \quad (9)$$

where p_i^h, p_i^s, p_i^{hyd}, p_i^c are the amino acid scales for α-helix and β-sheet formation, hydrophobicity and charge. An aggregation propensity profile is then defined as[25]

$$P_i^{agg} = \frac{1}{7}\sum_{j=-3}^{3} p_{i+j}^{agg} + \alpha_{pat} I_i^{pat} + \alpha_{gk} I_i^{gk} \quad (10)$$

where we considered the aggregation rate of a seven-residue segment of the protein centered at position i. I_i^{pat} and I_i^{gk} are included, respectively, to account for the presence of hydrophobic patterns and of gatekeeper residues.[25] The term I_i^{pat} is 1 if residue i is included in a hydrophobic pattern over five

residues and 0 otherwise, while the term I_i^{gk} is defined as[11,25]

$$I_i^{gk} = \sum_{j=-10}^{10} c_{i+j} \quad (11)$$

where the sum over the charges c_i of individual amino acids is made over a sliding window of 21 residues; shorter windows are considered at the N- and C-termini. The term I_i^{gk} is introduced to take into account the fact that when a hydrophobic pattern is flanked by charged residues its contribution to the aggregation propensity is much reduced by electrostatic repulsions.

By normalizing P_i^{agg}, the intrinsic aggregation propensity score, Z_i^{agg}, is obtained, which is used to compare the aggregation propensity of different sequences

$$Z_i^{agg} = \frac{P_i^{agg} - \mu^{agg}}{\sigma^{agg}} \quad (12)$$

where μ^{agg} is the average value of Z_i^{agg} over a set of random polypeptides having the same length as the sequence of interest, and σ^{agg} is the corresponding standard deviation from the average.[25] The corresponding intrinsic propensity for aggregation of the protein without structural correction, Z^{agg}, is calculated by adding the contributions from all residues with $Z_i^{agg} > 0$ and dividing by their number.

The local stability of various regions in a protein can be investigated by using the CamP method, which uses the knowledge of the amino acid sequence to provide a prediction of the protection factors from hydrogen exchange, ln P_i. By combining the predictions of the intrinsic aggregation propensity profiles with those for folding into stable structures, we account for the influence of the structural context on the aggregation propensities.[52] We thus define[25] a sequence-dependent aggregation propensity score with the structural correction, $\tilde{Z}_i^{agg}$, by modulating the intrinsic aggregation propensity profile with the local stability score:

$$\tilde{Z}_i^{agg} = Z_i^{agg}(1 - \varepsilon \ln P_i) \quad (13)$$

where ε was fixed at 1/15. These modulations on the aggregation-propensity profile are made only when $Z_i^{agg} > 0$ since we consider only the effects on the regions of high intrinsic aggregation propensity, which are those that effectively drive the aggregation process. The plot of $\tilde{Z}_i^{agg}$ *versus* the residue number represents the aggregation propensity profile calculated to account for the structural protection. The corresponding propensity for aggregation of the protein, $\tilde{Z}^{agg}$, in this case is calculated by adding the contributions from all residues $\tilde{Z}_i^{agg}$ and dividing by their total number N. Only residues with high $\tilde{Z}_i^{agg}$ values ($\tilde{Z}_{threshold}^{agg} = 0$) are considered as contributing to the aggregation propensity, $\tilde{Z}^{agg}$, resulting in the formula

$$\tilde{Z}^{agg} = \frac{\sum_{i=1}^{N} \tilde{Z}_i^{agg} \vartheta(\tilde{Z}_i^{agg})}{\sum_{i=1}^{N} \vartheta(\tilde{Z}_i^{agg})} \quad (14)$$

where the function $\vartheta(\tilde{Z}_i^{agg})$ is 1 for $(\tilde{Z}_i^{agg}) \geq 0$ and 0 for $(\tilde{Z}_i^{agg}) < 0$.

Conclusion

In this work we have shown that monellin contains two aggregation-prone segments that are protected from exposure to the solvent and aggregation by multiple non-covalent interactions that favor the folded state over the unfolded state of the protein. Weakening of these interactions by mutation leads to enhanced aggregation and formation of amyloid fibrils. The lag time for fibrillation correlates both with the apparent midpoint of thermal denaturation for a series of five mutants and with the predicted aggregation propensity when taking into account structural protection in the folded state. This study thus provides a detailed characterization of the interactions that help maintain a protein in its soluble state despite the presence of sequence regions of high aggregation propensity.

Abbreviations

CD	circular dichroism
CM	carboxymethyl
DEAE	diethylaminoethyl
IPTG	isopropyl-β-D-1-thiogalactoside
MES	2-(*N*-morpholino) ethanesulfonic acid
MN	wild type monellin composed of the A and B chains
MNA	monellin A chain
MNB	monellin B chain
scMN	single chain monellin
OD	optical density
aa	amino acid
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
ThT	thioflavin
T	Tris, tris(hydroxymethyl)aminomethane

Acknowledgements

This study has been supported by The Research School of Pharmaceutical Sciences at Lund University (FLÄK) and the Swedish Research Council. C.C.-L. acknowledges the *Isidro Parga Pondal* Program fellowship (Xunta de Galicia, Spain).

References

1 D. M. Fowler, A. V. Koulov, W. E. Balch and J. W. Kelly, *Trends Biochem. Sci.*, 2007, **32**, 217–224.
2 S. K. Maji, M. H. Perrin, M. R. Sawaya, S. Jessberger, K. Vadodaria, R. A. Rissman, P. S. Singru, K. P. Nilsson, R. Simon, D. Schubert, D. Eisenberg, J. Rivier, P. Sawchenko, W. Vale and R. Riek, *Science*, 2009, **325**, 328–332.
3 C. P. Maury, *J. Intern. Med.*, 2009, **265**, 329–334.
4 F. Chiti and C. M. Dobson, *Annu. Rev. Biochem.*, 2006, **75**, 333–366.
5 C. Haass and D. J. Selkoe, *Nat. Rev. Mol. Cell Biol.*, 2007, **8**, 101–112.
6 J. W. Kelly, *Structure*, 1997, **5**, 595–600.
7 R. Nelson, M. R. Sawaya, M. Balbirnie, A. Ø. Madsen, C. Riekel, R. Grothe and D. Eisenberg, *Nature*, 2005, **435**, 773–778.
8 C. M. Dobson, *Philos. Trans. R. Soc. Lond. Biol. Sci.*, 2001, **356**, 133–145.
9 F. Chiti, M. Stefani, N. Taddei, G. Ramponi and C. M. Dobson, *Nature*, 2003, **424**, 805–808.

10 R. Schwartz, S. Istrail and J. King, *Protein Sci.*, 2001, **10**, 1023–1031.
11 B. M. Broome and M. H. Hecht, *J. Mol. Biol.*, 2000, **296**, 961–968.
12 F. Chiti, M. Calamai, N. Taddei, M. Stefani, G. Ramponi and C. M. Dobson, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 16419–16426.
13 J. P. Schmittschmitt and J. M. Scholtz, *Protein Sci.*, 2003, **12**, 2374–2378.
14 M. Vendruscolo and C. M. Dobson, *Nature*, 2007, **449**, 555.
15 S. B. Zimmerman and S. O. Trach, *J. Mol. Biol.*, 1991, **222**, 599–620.
16 R. J. Ellis, *Curr. Opin. Struct. Biol.*, 2001, **11**, 114–119.
17 R. J. Ellis and A. P. Minton, *Biol. Chem.*, 2006, **387**, 485–497.
18 C. M. Dobson, *Semin. Cell Dev. Biol.*, 2004, **15**, 3–16.
19 P. J. Muchowski, *Neuron*, 2002, **35**, 9–12.
20 F. U. Hartl and M. Hayer-Hartl, *Science*, 2002, **295**, 1852–1858.
21 R. J. Kaufman, D. Scheuner, M. Schröder, X. Shen, K. Lee, C. Y. Liu and S. M. Arnold, *Nat. Rev. Mol. Cell Biol.*, 2002, **3**, 411–421.
22 A. M. Morris and R. G. Finke, *Biophys. Chem.*, 2009, **140**, 9–15.
23 K. Klement, K. Wieligmann, J. Meinhardt, P. Hortschansky, W. Richter and M. Fändrich, *J. Mol. Biol.*, 2007, **373**, 1321–1333.
24 E. Hellstrand, B. Boland, D. M. Walsh and S. Linse, *ACS Chem. Neurosci.*, 2010, **1**, 13–18.
25 G. G. Tartaglia, A. P. Pawar, S. Campioni, C. M. Dobson, F. Chiti and M. Vendruscolo, *J. Mol. Biol.*, 2008, **380**, 425–436.
26 P. Bornstein and H. Sage, *Annu. Rev. Biochem.*, 1980, **49**, 957–1003.
27 M. S. Nissen, G. N. Kumar, B. Youn, D. B. Knowles, K. S. Lam, W. J. Ballinger, N. R. Knowles and C. Kang, *Plant Cell*, 2009, **21**, 861–875.
28 J. A. Morris and R. H. Cagan, *Biochim. Biophys. Acta*, 1972, **26**, 114–122.
29 J. A. Morris and R. H. Cagan, *Proc. Soc. Exp. Biol. Med.*, 1980, **164**, 351–354.
30 J. R. Somoza, F. Jiang, L. Tong, C. Kang, J. Cho and S. H. Kim, *J. Mol. Biol.*, 1993, **234**, 390–404.
31 T. Konno, K. Murata and K. Nagayama, *FEBS Lett.*, 1999, **454**, 122–126.
32 T. Konno, *Protein Sci.*, 2001, **10**, 2093–2101.
33 W. F. Xue, O. Szczepankiewicz, E. Thulin, S. Linse and J. Carey, *Biochim. Biophys. Acta*, 2009, **1794**, 410–420.
34 M. Kohmura, N. Nio and Y. Ariyoshi, *Agric. Biol. Chem.*, 1990, **54**, 2219–2224.
35 W. F. Xue, J. Carey J and S. Linse, *Proteins: Struct., Funct., Bioinf.*, 2004, **57**, 586–595.
36 Z. Bohak and S. L. Li, *Biochim. Biophys. Acta*, 1976, **427**, 153–170.
37 C. Ogata, M. Hatada, G. Tomlinson, W. C. Shin and S. H. Kim, *Nature*, 1987, **328**, 739–742.
38 R. Day, D. A. Beck, R. S. Armen and V. Daggett, *Protein Sci.*, 2003, **12**, 2150–2160.
39 S. H. Kim, C. H. Kang, R. Kim, J. M. Cho, Y. B. Lee and T. K. Lee, *Protein Eng.*, 1989, **2**, 571–575.
40 A. K. Patra and J. B. Udgaonkar, *J. Mol. Biol.*, 2009, **389**, 759–775.
41 Y. H. Sung, J. Shin, H. J. Chang, J. M. Cho and W. Lee, *J. Biol. Chem.*, 2001, **276**, 19624–19630.
42 Y. H. Sung, H. D. Hong, C. Cheong, J. H. Kim, J. M. Cho, Y. R. Kim and W. Lee, *J. Biol. Chem.*, 2001, **276**, 44229–44238.
43 W. F. Xue, O. Szczepankiewicz, M. C. Bauer, E. Thulin and S. Linse, *J. Mol. Biol.*, 2006, **358**, 1244–1255.
44 E. M. Baden, E. G. Randles, A. K. Aboagye, J. R. Thompson and M. Ramirez-Alvarado, *J. Biol. Chem.*, 2008, **283**, 30950–30956.
45 F. Chiti, N. Taddei, M. Bucciantini, P. White, G. Ramponi and C. M. Dobson, *EMBO J.*, 2000, **19**, 1441–1449.
46 S. Pechmann, E. D. Levy, G. G. Tartaglia and M. Vendruscolo, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 10159–10164.
47 O. S. Makin, E. Atkins, P. Sikorski, J. Johansson and L. C. Serpell, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 315–320.
48 K. E. Amrein, B. Takacs, M. Stieger, J. Molnos, N. A. Flint and P. Burn, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 1048–1052.
49 J. Sambrook, E. F. Fritsch and T. Maniatis, *Molecular Cloning: a laboratory manual,* Cold Spring Harbour, NY, 1989.
50 L. Nielsen, R. Khurana, A. Coats, S. Frokjaer, J. Brange, S. Vyas, V. N. Uversky and A. L. Fink, *Biochemistry*, 2001, **40**, 6036–6046.
51 M. Kodaka, *Biophys. Chem.*, 2004, **107**, 243–253.
52 G. G. Tartaglia, A. Cavalli and M. Vendruscolo, *Structure*, 2007, **5**, 139–143.
53 J. J. Balbach, Y. Ishii, O. N. Antzutkin, R. D. Leapman, N. W. Rizzo, F. Dyda, J. Reed and R. Tycko, *Biochemistry*, 2000, **39**, 13748–13759.
54 K. Tenidis, M. Waldner, J. Bernhagen, W. Fischle, M. Bergmann, M. Weber, M. L. Merkle, W. Voelter, H. Brunner and A. Kapurniotu, *J. Mol. Biol.*, 2000, **295**, 1055–1071.
55 O. S. Makin, E. Atkins, P. Sikorski, J. Johansson and L. C. Serpell, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 315–320.
56 W. Hosia, N. Bark, E. Liepinsh, A. Tjernberg, B. Persson, D. Hallén, J. Thyberg, J. Johansson and L. Tjernberg, *Biochemistry*, 2004, **43**, 4655–4661.
57 S. Jones, J. Manning, N. M. Kad and S. E. Radford, *J. Mol. Biol.*, 2003, **325**, 249–257.
58 G. T. Westermark, U. Engström and P. Westermark, *Biochem. Biophys. Res. Commun.*, 1992, **182**, 27–33.
59 J. Ghiso, R. Vidal, A. Rostagno, S. Mead, T. Révész, G. Plant and B. Frangione, *Ann. N. Y. Acad. Sci.*, 2000, **903**, 129–137.
60 M. Balbirnie, R. Grothe and D. S. Eisenberg, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 2375–2380.