

1-1-2008

Dissociation from the oligomeric state is the rate-limiting step in fibril formation by kappa-casein

Heath Ecroyd

University of Adelaide, heathe@uow.edu.au

Tomas Koudelka

University of Adelaide, tomas.koudelka@adelaide.edu.au

David Thorn

University of Adelaide

Danielle M. Williams

University of Adelaide, danielle_williams@adelaide.edu.au

Glyn Devlin

University of Cambridge

See next page for additional authors

Follow this and additional works at: <https://ro.uow.edu.au/scipapers>



Part of the [Life Sciences Commons](#), [Physical Sciences and Mathematics Commons](#), and the [Social and Behavioral Sciences Commons](#)

Recommended Citation

Ecroyd, Heath; Koudelka, Tomas; Thorn, David; Williams, Danielle M.; Devlin, Glyn; Hoffmann, Peter; and Carver, John A.: Dissociation from the oligomeric state is the rate-limiting step in fibril formation by kappa-casein 2008, 9012-9022.
<https://ro.uow.edu.au/scipapers/930>

Dissociation from the oligomeric state is the rate-limiting step in fibril formation by kappa-casein

Abstract

Amyloid fibrils are aggregated and precipitated forms of protein in which the protein exists in highly ordered, long, unbranching threadlike formations that are stable and resistant to degradation by proteases. Fibril formation is an ordered process that typically involves the unfolding of a protein to partially folded states that subsequently interact and aggregate through a nucleation-dependent mechanism. Here we report on studies investigating the molecular basis of the inherent propensity of the milk protein, kappa-casein, to form amyloid fibrils. Using reduced and carboxymethylated kappa-casein (RCM kappa-CN), we show that fibril formation is accompanied by a characteristic increase in thioflavin T fluorescence intensity, solution turbidity, and beta-sheet content of the protein. However, the lag phase of RCM kappa-CN fibril formation is independent of protein concentration, and the rate of fibril formation does not increase upon the addition of seeds (preformed fibrils). Therefore, its mechanism of fibril formation differs from the archetypal nucleation-dependent aggregation mechanism. By digestion with trypsin or proteinase K and identification by mass spectrometry, we have determined that the region from Tyr(25) to Lys(86) is incorporated into the core of the fibrils. We suggest that this region, which is predicted to be aggregation-prone, accounts for the amyloidogenic nature of kappa-casein. Based on these data, we propose that fibril formation by RCM kappa-CN occurs through a novel mechanism whereby the rate-limiting step is the dissociation of an amyloidogenic precursor from an oligomeric state rather than the formation of stable nuclei, as has been described for most other fibril-forming systems.

Keywords

rate, limiting, step, formation, kappa, state, oligomeric, fibril, casein, dissociation, CMMB

Disciplines

Life Sciences | Physical Sciences and Mathematics | Social and Behavioral Sciences

Publication Details

Ecroyd, H., Koudelka, T., Thorn, D., Williams, D. M., Devlin, G., Hoffmann, P. & Carver, J. A. (2008).

Dissociation from the oligomeric state is the rate-limiting step in fibril formation by kappa-casein. *Journal of Biological Chemistry*, 283 (14), 9012-9022.

Authors

Heath Ecroyd, Tomas Koudelka, David Thorn, Danielle M. Williams, Glyn Devlin, Peter Hoffmann, and John A. Carver

DISSOCIATION FROM THE OLIGOMERIC STATE IS THE RATE-LIMITING STEP IN FIBRIL FORMATION BY κ -CASEIN*

Heath Ecroyd¹, Tomas Koudelka^{1,2}, David C. Thorn¹, Danielle M. Williams¹, Glyn Devlin³, Peter Hoffmann², John A. Carver¹.

¹School of Chemistry and Physics and ²School of Molecular and Biomedical Science, University of Adelaide, Adelaide, SA, 5005 Australia and ³Department of Chemistry, University of Cambridge, Cambridge, CB2 1EW, U. K.

Running head: Amyloid fibril formation by κ -casein

Address correspondence to: Heath Ecroyd, School of Chemistry and Physics, The University of Adelaide, Adelaide, SA 5005, Australia, Tel. +61 8 830 35505; Fax. +61 8 830 34358; E-mail: heath.ecroyd@adelaide.edu.au

Amyloid fibrils are aggregated and precipitated forms of protein in which the protein exists in highly ordered, long, unbranching thread-like formations which are stable and resistant to degradation by proteases. Fibril formation is an ordered process that typically involves the unfolding of a protein to partially folded states that subsequently interact and aggregate through a nucleation-dependent mechanism. Here we report on studies investigating the molecular basis of the inherent propensity of the milk protein, κ -casein, to form amyloid fibrils. Using reduced and carboxymethylated κ -casein (RCM κ -CN) we show that fibril formation is accompanied by a characteristic increase in thioflavin T fluorescence intensity, solution turbidity and β -sheet content of the protein. However, the lag-phase of RCM κ -CN's fibril formation is independent of protein concentration and the rate of fibril formation does not increase upon addition of seeds (pre-formed fibrils). Therefore, its mechanism of fibril formation differs from the archetypal nucleation-dependent aggregation mechanism. By digestion with trypsin or proteinase K and identification by mass spectrometry we have determined that the region from Tyr²⁵ to Lys⁸⁶ is incorporated into the core of the fibrils. We suggest that this region, which is predicted to be aggregation prone, accounts for the amyloidogenic nature of κ -casein. Based on these data, we propose that fibril formation by RCM κ -CN occurs through a novel mechanism whereby the rate-limiting step is the dissociation of an amyloidogenic precursor from an oligomeric state, rather than the formation of stable nuclei as has been described for most other fibril-forming systems.

Amyloid fibrils are ordered aggregates of a normally soluble peptide or protein. Their formation results from the ordered association of partially folded protein intermediates that have a propensity to adopt a β -sheet conformation. These intermediates expose greater hydrophobicity to solution which encourages their mutual association. Fibril nucleation and elongation may then occur through sequential addition of intermediates due to β -sheet stacking via a nucleation-dependent mechanism (1-3). Typically the fibrils formed are long, straight, thread-like structures that are unbranched, 6-12 nm in diameter and may consist of two to six 'protofilaments' twisted around a central core. They are often rigid, stable and resistant to degradation (e.g. from proteases and denaturants).

Amyloid fibril formation is often associated with disease and the misfolding process is believed to be causative, or at least linked, to disease onset and progression (e.g. in Alzheimer's and Parkinson's diseases and various amyloidoses) (reviewed in (2,4)). However, amyloid fibril formation is not confined to disease-related proteins. Studies have shown that non-disease related proteins, even those that have a high α -helix content in their native state (5-8), can assemble into amyloid fibrils and this has led to the proposal that the amyloid state is a generic conformation accessible to all polypeptide chains under appropriate conditions (9). Whilst it may be true that all proteins are capable of forming fibrils given the right conditions, the composition and amino acid sequence of a protein profoundly affects its propensity to adopt such structures. Moreover, the tendency of a polypeptide chain to aggregate, rather than fold correctly, depends on a number of intrinsic factors including the

propensity for it to form β -strands, its hydrophobicity and overall net charge (10,11).

It has been previously shown that the reduced and carboxymethylated form of the milk protein, κ -casein (RCM κ -CN), forms amyloid fibrils when incubated under conditions of physiological pH and temperature (12,13). Furthermore, we have shown that native (non-reduced) κ -casein is also capable of forming fibrils under the same conditions (13) which is not surprising since reduction and carboxymethylation of the protein does not significantly alter its secondary or tertiary structure (12). From our previous studies, we concluded that κ -casein has an inherent susceptibility toward fibril formation, however, under physiological conditions in milk, this is kept in check by the action of the other caseins (α_s and β), which are able to inhibit its fibrillation (13). Of physiological relevance, amyloid-like deposits have been found in bovine mammary glands within calcified stones known as *corpora amylacea* (14-16) and bundles of fibrils have also been reported in the cytoplasm of cells that surround these calcified deposits (17). These fibrils may result from the incapacity of other casein proteins to prevent amyloid formation by κ -casein.

Casein is the principal component of milk micelles and is a heterogeneous mix of proteins that include α_{S1} -, α_{S2} -, β - and κ -casein. These proteins have been classified as intrinsically disordered (18,19) as they are extremely flexible, essentially unfolded and have relatively little secondary or tertiary structure under physiological conditions. Although the caseins themselves are relatively small, ranging in mass between 19 to 25 kDa, they are usually associated, with calcium and other ions, in milk micelles that have molecular masses from 1×10^3 to 3×10^6 kDa (20). When purified from the other casein proteins, κ -casein in aqueous solution forms large spherical polymers with an average molecular mass of 1.18 MDa (21). These polymers are formed from smaller multimeric subunits (ranging from monomers to decamers) that result from intramolecular and intermolecular disulfide bonding (22). When completely reduced, κ -casein forms a self-associating polymeric species (comprising 31 monomers) in equilibrium with the monomeric form (23). It has been proposed that the monomeric form of the protein is a precursor for amyloid fibril formation (13).

The inherent propensity for κ -casein to form amyloid fibrils may be reflective of the relatively unstructured nature of its monomeric form in solution. A theoretical energy minimized three-dimensional model of κ -casein proposed by Kumosinski *et al.* (24) suggests that the monomeric form of the protein can adopt a 'horse and rider' configuration in which the legs of the horse are two sets of anti-parallel β -sheets (from Lys²¹ to Phe⁵⁵), which are rich in hydrophobic side chains. It has been speculated that the putative anti-parallel β -sheet legs, due to their high degree of hydrophobicity and exposure to solution, may facilitate amyloid fibril formation by κ -casein (12).

The capacity of RCM κ -CN to form fibrils in a short time frame under conditions of physiological temperature and pH and without the need for denaturants, make it an ideal system in which to study amyloid fibril formation. Moreover, the link between casein amyloid fibril formation, *corpora amylacea* and the intracellular fibril-like bundles identified in cells surrounding these deposits indicates that a more detailed understanding of the mechanism by which κ -casein forms fibrils is required. Here, we show that fibril formation by RCM κ -CN occurs through a unique mechanism whereby the rate-limiting step is the dissociation of an amyloidogenic precursor from an oligomeric state. Furthermore, we have identified the region of the protein incorporated into the core of the amyloid fibrils formed by RCM κ -CN and show that this region encompasses that predicted to adopt an anti-parallel β -sheet structure in the native form of the protein (24) thereby providing the basis behind its unique fibril-forming mechanism. These results help to explain how fibril formation by κ -casein is largely prevented *in vivo*. Similar mechanisms involving dissociation of an amyloidogenic precursor from its natural binding partner/oligomeric state are also likely to be important in other fibril-forming systems *in vivo*.

Experimental Procedures

Materials- Bovine milk κ -casein (SWISS PROT accession number P02668), without the N-terminal 21 amino acid leader peptide, was purchased from Sigma Chemical Co. (St. Louis, MO, USA). The homogeneity of this protein was typically above 90% as assessed by gel electrophoresis and ion-exchange

chromatography. Further purification of the protein by ion-exchange chromatography showed that there was no change in the rate or degree of fibril formation in this sample compared to the protein supplied directly from Sigma (data not shown), indicating that the other minor components present did not influence the ability of reduced κ -casein to form fibrils. As such, we used the κ -casein without further purification in these studies. For the experiments described here the protein was reduced and carboxymethylated as described previously (12). The vector pRSETB (Invitrogen, Carlsbad, CA, U.S.A) containing the human α -synucleinA53T gene was a kind gift from Dr Roberto Cappai (University of Melbourne, Australia) and the α -synucleinA53T protein was expressed and purified as described previously (25). Thioflavin T (ThT), 1,4-dithiothreitol (DTT), proteinase K (molecular biology grade) and β -mercaptoethanol were obtained from Sigma. Trypsin (proteomics grade) was also purchased from Sigma and prepared at 1 mg/mL in HCl. Uranyl acetate was purchased from Agar Scientific (Stansted, UK). The concentration of κ -casein was determined by spectrophotometric methods using a Cary 5000 UV-Vis-NIR spectrophotometer (Varian, Melbourne, Australia) and an extinction coefficient of 0.95 mL.mg⁻¹.cm⁻¹ (26).

Thioflavin T and light scattering assays - The formation of amyloid fibrils by RCM κ -CN was monitored using an *in situ* ThT binding assay (27). The protein was incubated at 37 °C in 50 mM phosphate buffer, pH 7.2, unless stated otherwise. Samples were incubated with 10 μ M ThT in black μ Clear 96-microwell plates (Greiner Bio-One, Stonehouse, UK). The plates were sealed to prevent evaporation and the fluorescence levels were measured with a Fluostar Optima plate reader (BMG Labtechnologies, Melbourne, Australia) with a 440/490 nm excitation/emission filter set. Fibril formation was also monitored by light scattering at 340 nm using a Fluostar Optima plate reader. For each assay, samples were prepared in duplicate and the change in ThT fluorescence at 490 nm or light scattering at 340 nm is presented in the figures.

In the seeding experiments, RCM κ -CN (0.2 - 2.5 mg/mL) was incubated under the same conditions as described above, however, prior to the start of the incubation an aliquot of freshly prepared fibrils was added to the reaction such

that both the amount and the concentration of RCM κ -CN in each sample was maintained. In some experiments the fibrils were sonicated for 10 min prior to use. As a control for these experiments, α -synucleinA53T was incubated at 37°C in 50 mM phosphate buffer containing 100 mM NaCl, pH 7.4 and 10 μ M ThT in a microtitre plate. Single timepoint ThT readings were taken for α -synucleinA53T during incubation and the microtitre plate was subjected to constant shaking (700 rpm) using a ThermoStar microplate shaker (BMG Labtechnologies) between readings.

Transmission electron microscopy (TEM) - Formvar and carbon-coated nickel electron microscopy grids (SPI supplies, Wester Chester, USA) were prepared by the addition of 2 μ L of protein sample, washed with 3 x 10 μ L of milli-Q and negatively stained with 10 μ L of uranyl acetate (2% w/v). Samples were viewed using a Philips CM100 transmission electron microscope (Philips, Eindhoven, Netherlands) at a magnification range of 10,500 - 96,000 using an 80 kV excitation voltage.

Circular dichroism - Circular dichroism (CD) spectra over a wavelength range of 190 nm to 250 nm were acquired for samples of RCM κ -CN before and after fibril formation using a Jasco J-180 spectropolarimeter (Jasco, Victoria, Canada) connected to a Jasco water bath using a cell with a 0.1 mm optical pathlength. Samples were prepared in 20 mM phosphate buffer, pH 7.2 and diluted with the same buffer to a final concentration of 0.2 mg/mL before obtaining the spectra. The spectra were obtained at 37°C. Solvent spectra were subtracted from the measured protein spectra and the results shown are the average of 5 scans.

FTIR Spectra and Deconvolution Procedures - Samples of RCM κ -CN before and after fibril formation were analyzed in a Bruker BioATRCCell II using a Bruker Equinox 55 Fourier transform infrared spectrometer (MA, USA) equipped with a liquid nitrogen cooled mercury cadmium telluride detector and a silicon internal reflection element. For each spectrum, 256 interferograms were co-added at 2 cm⁻¹ resolution, and the water background was independently measured and subtracted from each protein spectrum. Spectral contributions from water vapor were subtracted using the atmospheric compensation feature of the Bruker Opus software package. The amide I region of the IR spectrum (1600 – 1700 cm⁻¹) consists of a

series of overlapped component bands which arise from absorption by different secondary structural elements in proteins and peptides. In order to deconvolve these spectra, second derivatives of the amide I region were taken to determine the peak positions of the component bands. Subsequently, these component bands within the original spectra were fitted to a fixed number of Gaussian functions (28) using the OriginPro 7.5 software package by OriginLab (MA, USA), where the heights, widths and positions of each band were optimized iteratively.

Proteolytic cleavage of amyloid fibrils formed by RCM κ -CN - Prior to proteolysis, RCM κ -CN fibrils were formed by incubating in 25 mM ammonium bicarbonate buffer, pH 7.4, overnight at 37°C. The presence of fibrils was confirmed by TEM. Limited proteolysis experiments were carried out by incubating native and fibrillar forms of RCM κ -CN with trypsin or proteinase K. Enzymatic digestions were performed at 37°C using enzyme-to-substrate ratios of 1: 100 (w/w) for trypsin and 1: 1000 (w/w) for proteinase K. The presence of amyloid fibrils following proteolysis was confirmed by TEM. At the end of the assay an aliquot of each sample was taken for SDS-PAGE analysis and the remainder used for MALDI-TOF MS analysis. For SDS-PAGE, aliquots were removed at various time intervals and the reaction stopped by mixing with an equal volume of reducing sample buffer (50 mM Tris-HCl, pH 6.8, 4% w/v SDS, 2% v/v 2-mercaptoethanol, 12% glycerol, 0.01% bromophenol blue). These samples were then heated to 95°C for 5 min before being loaded onto gels. Proteolysis products were analyzed using 15% or 20% acrylamide gels (v/v) and standard techniques (29). Analysis of the samples by TEM showed that the sample preparation for SDS-PAGE analysis was sufficient to disaggregate the fibrils.

Mass spectrometry - An Ultraflex III MALDI-TOF/TOF mass spectrometer (Bruker Daltonics, Bremen, Germany) was used for the analysis of intact protein and proteolytic cleaved peptides. The instrument was used in linear mode for peptides >6000 Da and in reflectron mode for smaller peptides. Linear mode was calibrated using protein calibration standard I (Bruker Daltonics). The protein mixture enables calibrations and testing in a mass range of approximately 3000 to 25000 Da with an error of approximately 100 ppm. Reflectron mode was

calibrated using peptide calibration standard II (Bruker Daltonics) which enables calibrations and testing in a mass range 500- 6000 Da to a mass accuracy greater than 50 ppm.

Aliquots taken for analysis by mass spectrometry were firstly mixed with urea to a final concentration of 6 M and left overnight at room temperature in order to disaggregate the fibrils. The sample was then mixed with trifluoroacetic acid (TFA) to a final concentration of 0.1% (v/v), acidified with HCl (final pH < 4), before being taken up in a reverse phase column ZipTip_{C-18} (Millipore, Billerica, MA, USA). After washing with acetonitrile (5% v/v), containing 0.1% (v/v) TFA multiple times to remove any salt and urea present, the sample was then step-wise eluted with increasing concentrations of acetonitrile (20%, 30%, 40%, 50% and 60% v/v in 0.1% v/v TFA). The proteinase resistant peptide of interest was eluted with 50% and 60% v/v acetonitrile. Samples were analyzed by MALDI-TOF/TOF MS in linear mode using 2,5-dihydroxybenzoic acid (2,5-DHB) as the matrix (5 g/L) in 50% v/v acetonitrile with 0.1% v/v TFA.

To confirm this peptide matched the protease-resistant band identified by SDS-PAGE, a tryptic in-gel digest was performed according to the method of Shevchenko (30) with the following modifications. The Coomassie blue-stained bands from the gel were cut with a sterile scalpel blade into small blocks and destained using 50 mM NH₄HCO₃ in 30% v/v acetonitrile. The blocks were rinsed and incubated overnight at 37°C with 10 ng/ μ L of trypsin in 5 mM NH₄HCO₃. The tryptic fragments were extracted from the gel pieces by sonicating the samples for 15 min with 50% v/v acetonitrile with 0.1% v/v TFA and then the extracted peptide solution was dried down to a final volume of approximately 5 μ L using a vacuum centrifuge. The tryptic peptides were analyzed directly by MALDI-TOF MS in reflector mode using 4-HCCA (α -Cyano-4-hydroxycinnamic acid) as the matrix. The identity of these peptides was confirmed by matching their observed masses to the expected masses of the *in silico* tryptic digest of the known κ -casein sequence using Biotoools 3.0 software (Bruker Daltonics).

Results

Amyloid fibril formation by RCMκ-CN does not proceed by a standard nucleation-dependent mechanism. Firstly, RCMκ-CN was incubated at 37°C at various concentrations (0 – 10 mg/mL) in 50 mM phosphate buffer, pH 7.2 and the increase in ThT binding (which is associated with fibril formation) was examined by an *in situ* assay (Fig 1A and B). We have previously observed that the presence of the ThT dye in this assay does not affect the rate or extent of fibril formation by RCMκ-CN or the morphology of the fibrils formed. For all the concentrations tested, we found that during the early phase of incubation (i.e. the first 30 min), there was an initial concentration-dependent decrease in the ThT fluorescence of the samples. We and others have shown that κ-casein is able to bind ThT in its native state (12,13) and we observed that this initial decrease in fluorescence intensity at the start of this assay was attributable to the temperature-dependence of the ThT fluorescence (an increase in temperature leading to a decrease in fluorescence intensity). Thus, as the samples were warmed from room temperature to 37°C the fluorescence intensity from ThT bound to the native state of the protein decreased. After this period, there was a concentration-dependent increase in ThT fluorescence which reached a maximum at 600 – 800 min. The presence of fibrils in these samples was confirmed by TEM. Since the rate of increase in ThT fluorescence was observed to be dependent on the protein concentration (Fig. 1A and B), we sought to determine the reaction rate for the formation of fibrils by RCMκ-CN. The apparent reaction order can be determined according to the equation $\ln V = A + n \ln C$, where V is the initial reaction rate, C is the protein concentration and n is the apparent reaction order. The plot of the log of the protein concentration against the log of the initial reaction rate (Fig 1C), gave a slope (n) of 1.13 ± 0.15 , indicating that the rate of increase in ThT fluorescence for RCMκ-CN follows an apparent first-order kinetic mechanism in relation to the protein concentration.

We also monitored the incubation of RCMκ-CN by a light scattering assay, which did not involve ThT (Fig. 1D). When RCMκ-CN (0 - 10 mg/mL) was incubated under the same conditions as for the ThT assay there was a concentration-dependent increase in light

scattering that was associated with fibril formation. Interestingly, both the ThT *in situ* (Fig. 1A and B) and light scattering (Fig. 1D) data showed no evidence of a concentration-dependent variation in the lag phase associated with fibril formation, an attribute that is a component of a nucleation-dependent mechanism of aggregation (such as has been described for most amyloid fibril-forming systems) (3). Even when the protein was incubated at lower temperatures, e.g. 32°C (Fig. 1E) and 27°C (data not shown), the lag phase of the increase in ThT fluorescence was very short (i.e. < 30 min) and concentration independent. According to a nucleation-dependent aggregation mechanism, the rate-limiting step of fibril formation is the generation of stable nuclei, the rate of which is limited by the protein concentration. As such, the length of the lag phase of conversion of the protein to amyloid fibrils depends dramatically on the concentration of the protein (3). The clear absence of a concentration-dependent lag phase associated with fibril formation by RCMκ-CN suggested that the mechanism by which it forms fibrils may differ from that typically reported for other systems.

We also tested whether adding preformed fibrils (seeds) to RCMκ-CN (2.5 mg/mL) incubated at 37°C increased the rate of fibril formation, since in a nucleation-dependent mechanism, addition of exogenous seeds bypasses the need for nucleus formation, which, as discussed above, is typically considered to be the rate-limiting step in the aggregation process. However, addition of seeds (0.5 – 10.0% w/w) had no effect on the length of the lag phase or the rate of increase in ThT fluorescence, either at 32°C (data not shown) or 37°C (Fig. 1F). Pre-formed fibrils that were sonicated prior to addition also had no effect on RCMκ-CN fibril formation (data not shown). Similar results were found when the protein was incubated at 37°C at 0.2 mg/mL and the reaction was seeded (10.0% w/w) with preformed fibrils (data not shown). In contrast, seeding (0.5 – 2.0% w/w) a sample of α-synucleinA53T (2 mg/mL), the Parkinson's disease related fibril-forming protein which has been shown to form fibrils through a classic nucleation-dependent mechanism, dramatically increased the rate of fibril formation and decreased the lag phase of the reaction (Fig. 1F – inset). Together, these data suggest that the rate of fibril formation by RCMκ-CN is not limited

by the formation of stable nuclei as is found in most nucleation-dependent mechanisms of aggregation, but, instead is dependent on steps that precede the formation of nuclei, for example the dissociation of an amyloidogenic precursor from its oligomeric state.

Monitoring the changes in RCMκ-CN before and after fibril formation by spectroscopic methods - Since RCMκ-CN has an inherent propensity to form fibrils when incubated under conditions of physiological pH and temperature, we investigated structural changes in RCMκ-CN that occur during the process of amyloid fibril formation to see if these would help to explain its amyloidogenic nature. RCMκ-CN (1 mg/mL) was incubated under conditions whereby it readily forms fibrils, i.e. 50 mM phosphate buffer, pH 7.2 at 37°C for 1000 min, and the presence of fibrils was confirmed by ThT fluorescence and TEM. Far-UV circular dichroism (CD) spectroscopy (Fig. 2A), used to probe the secondary structural changes that accompany RCMκ-CN fibril formation, showed that freshly prepared RCMκ-CN gives a spectrum with a maximum negative molar ellipticity at 200 nm and shoulder at 215 nm, indicative of a predominately random coil protein and consistent with CD spectra of the protein reported previously (12). Following incubation under fibril-forming conditions, the CD spectrum of RCMκ-CN was red shifted such that the maximum negative molar ellipticity was at 203 nm and was decreased in intensity, and the shoulder at 215 nm was increased indicating that upon fibril formation there is a loss in random coil and an increase in β -sheet structure compared to the native protein. Similarly, FTIR analysis of the protein before and after fibril formation showed an increased β -sheet content in the fibrillar form of the protein (Fig. 2C). The shift in the amide I maximum to lower wavenumbers (from 1645 cm^{-1} to 1635 cm^{-1}) suggests that there is significant structural reorganization during fibril formation involving an increase in β -sheet structure (31). From the deconvolved data (Fig. 2D), showing the individual component bands of the spectra (32), an increase in the band at ca. 1625 cm^{-1} and the development of bands at 1639 cm^{-1} and 1677 cm^{-1} following fibril formation is attributable to the increase in β -sheet content in the sample. There was a corresponding decrease in bands at ca. 1669 cm^{-1} and 1685 cm^{-1} attributed to turns and bends, while the magnitude of the band

corresponding to disordered structure (1648 - 1653 cm^{-1}) was largely unchanged. Overall, the deconvolved FTIR data indicated a ~2.4-fold increase in the amount of β -sheet in RCMκ-CN following fibril assembly (from 17 to 41%) which is primarily derived from the loss of turns and bends present in the non-fibrillar form of the protein (Table 1).

Identification of the protease-resistant core region of fibrils formed from RCMκ-CN - Our data suggest that the critical, rate-limiting step in amyloid fibril formation by RCMκ-CN is the dissociation of a precursor that is highly amyloidogenic and readily incorporated into pre-fibrillar nuclei. Thus, we sought to identify the region of the protein incorporated into the core of the amyloid fibrils formed by RCMκ-CN in order to investigate whether this region showed characteristics of being highly amyloidogenic. In order to identify the core region of these fibrils we performed proteolysis experiments on native (N) and fibrillar (F) forms of the protein. In the absence of trypsin, RCMκ-CN derived from fibrils was observed as a full-length protein (19 kDa) by SDS-PAGE. Addition of trypsin (1:100 w/w ratio of substrate: trypsin) to native RCMκ-CN resulted in its rapid proteolysis in less than 30 min, generating small peptide fragments of less than 1 kDa (Fig. 3A). In contrast, fibrillar RCMκ-CN was much more resistant to degradation, such that, even following 4 hr of incubation with trypsin, a protease-resistant peptide of ~7 kDa was still evident by SDS-PAGE. Following proteolysis with trypsin, fibrils with a similar morphology to those present before digestion were observed by TEM (Fig. 3B - C).

Similar results were obtained when proteinase K was used in place of trypsin (Fig. 4A), i.e. native RCMκ-CN was degraded by proteinase K into small peptide fragments whereas the fibrillar form was resistant to proteolytic degradation. After 4 hr of incubation of fibrillar with RCMκ-CN with proteinase K, four peptide bands were observed by SDS-PAGE at 18, 15, 10 and 8 kDa. Following digestion with proteinase K, amyloid fibrils were still observed by TEM (Fig. 4B-C).

We identified the protease-resistant core region of these amyloid fibrils by mass spectrometry. Since digestion with trypsin produced a smaller protease-resistant peptide fragment compared to proteinase K, we used this treatment in order to identify the protease-resistant core. Following disaggregation of the

trypsin-treated fibrils, we purified the resultant peptide and subjected it to MALDI-TOF MS. The peptide was seen as a $[M+H]^+$ ion at 7207.66 Da in the spectrum (Fig. 5A) and thus its mass matched that of the peptide identified by SDS-PAGE (see Fig. 3A). A number of smaller tryptic fragments were also identified that matched fragments generated when the native form of the protein was treated with trypsin (data not shown). Based on its mass, the identity of this tryptic fragment could not be unambiguously assigned: it could correspond to fragment κ -casein₂₅₋₈₆ or κ -casein₃₅₋₉₇ (Table 2). We confirmed the identity of the peptide by performing in-gel tryptic digestion of the band identified by SDS-PAGE and analyzing the fragments by MALDI-MS/MS. Three peaks with m/z ratios of 4009.90, 1980.03 and 1251.70 (Fig 5B) were identified that matched to tryptic fragments of κ -casein₃₅₋₆₈, κ -casein₆₉₋₈₆, and κ -casein₂₅₋₃₄ respectively (Table 2), and thus correspond to fragments of the κ -casein₂₅₋₈₆ peptide. The sequence identity of these in-gel tryptic fragments was confirmed by tandem mass spectrometry. The peaks with lower m/z ratios in the spectrum arose from further fragmentation of these peptide fragments.

We also confirmed that the same region of the protein was resistant to proteolytic degradation by proteinase K by performing an in-gel digest of the bands identified by SDS-PAGE (see Fig. 4A) and analyzing the fragments by MALDI-MS/MS (data not shown). Fragments corresponding to κ -casein₁₇₋₁₁₁ and κ -casein₁₇₋₈₆ were identified from the 10 kDa and 8 kDa bands respectively (Table 2). Both of these fragments encompass the region identified by trypsin proteolysis of the fibrils. Similar tryptic fragments were identified from the 18 kDa and 15 kDa bands and therefore these higher molecular mass bands incorporate the same protease-resistant region of κ -casein and represent larger peptides that were not fully digested by the proteinase K treatment.

Correlation of the protease-resistant core region of fibrils formed by RCM κ -CN with theoretical predictions on the structure and fibril-forming propensity of κ -casein - Whilst κ -casein is generally classified as an intrinsically disordered protein (18,19), previous three-dimensional modelling (24) indicates that it can sample defined conformations, a finding that is supported by our CD and FTIR data (see Fig. 2). We have employed two independent disorder-predicting algorithms to assess the regions of the

κ -casein that are likely to fold into a well-defined conformation, and therefore facilitate fibril formation (Fig. 6). Analysis with the VLXT algorithm using PONDR (33,34) indicated that, whilst almost half of the residues in the protein are predicted to be disordered (44%), the region Lys²⁴-Pro⁶⁴ is the most likely to adopt an ordered conformation. A similar region of the protein (Tyr³⁰-Gln⁸⁹) was predicted to fold into a defined conformation using the FoldIndex algorithm (35,36). Both these regions of predicted order therefore correlate very well with the region identified herein as being incorporated into the fibrillar core of fibrils formed by RCM κ -CN (Tyr²⁵-Lys⁸⁶) suggesting that this region is likely to adopt the β -sheet conformation predicted by the previous modelling work (24).

An examination of so-called “amyloidogenic domains” within a number of amyloid-forming proteins has led to the development of algorithms to predict aggregation-prone regions. We have used two such algorithms to examine the regions of κ -casein predicted to be the most aggregation prone; (1) the algorithm TANGO (37) indicates these to be residues Ile²⁸-Leu³² and Ile⁷³-Leu⁷⁹ and, (2) a modified version of the Zyggregator algorithm (38) (Tartaglia *et al.*, in preparation), also predicts regions of significant aggregation propensity between Tyr²⁵-Gln²⁹ and Leu⁷⁴-Ser⁸⁰ (Fig. 6). κ -Casein also contains two regions (Tyr³⁸-Gln⁴⁴ and Thr¹²¹-Ala¹²⁶) that match the six amino acid sequence pattern proposed by Lopez de la Paz and Serrano (39) to identify amyloidogenic stretches within proteins, however, the latter is preceded by a proline residue at position 120 which was found to be an “amyloid breaker”. Hence, these predictive methods correlate well with the protease-resistant fragment (Tyr²⁵-Lys⁸⁶) identified in this study, which includes at least two regions of high aggregation propensity, encompassing residues Tyr²⁵-Gln⁴⁴ and Ile⁷³-Ser⁸⁰.

Discussion

Fibril formation by RCM κ -CN does not follow a simple nucleation-dependent mechanism - We have previously shown that in its native state κ -casein has an inherent propensity to form amyloid fibrils when incubated under conditions of physiological pH and temperature *in vitro* (13). The formation of fibrils is enhanced by reduction of its disulfide bonds, however, this is

not attributable to significant secondary or tertiary structural changes in the protein upon reduction (12). As a result, we have used RCM κ -CN in these studies since it forms fibrils much more readily than the non-reduced 'native' form (13) and we attribute this primarily to the absence of disulfide bonds, thereby increasing the pool of dissociated species in solution that is available to form fibrils.

The key to understanding the rate of amyloid formation involves the identification of the rate-determining step in the overall process. Amyloid fibril formation is generally considered to proceed via a nucleation-dependent mechanism in which the slow step is the formation of an ordered, stable nucleus (Fig. 7) (3). The time required for the formation of this nucleus corresponds to a lag phase of polymerization, which is strongly protein concentration-dependent. In this scheme, the lag phase can be shortened by addition of preformed nuclei (seeds) to the reaction mix, which overcomes the time required for the formation of these seeds. Since fibril elongation also occurs from these added nuclei, the rate of fibril formation is also increased by seeding in a simple nucleation-dependent model (with the increase in rate being proportional to the amount of seeds added). In this study we have shown that RCM κ -CN fibril formation appears to be regulated by a novel mechanism, i.e. not the formation of a stable nucleus as is the case in a simple nucleation-dependent model. We have three distinct lines of evidence that have led us to this conclusion. Firstly, the lag phase of fibril formation is independent of the protein concentration (Fig. 1A, B and D) suggesting that steps preceding the formation of a nucleus are significant in the reaction mechanism. Secondly, this preceding step is rate-limiting and follows an approximate first-order reaction mechanism (Fig. 1C) with respect to the protein concentration. Thirdly, addition of exogenous seeds to the reaction did not affect either the rate or lag phase of fibril formation (Fig. 1F).

Together these data are consistent with a model in which the dissociated form of the protein is the precursor to fibril formation. Moreover, the dissociation of this amyloidogenic species from an oligomeric state is the rate-limiting step to fibril formation by RCM κ -CN (Fig. 7). This dissociation step therefore limits the supply of species to form stable nuclei and the supply of precursor to the growing fibril during its elongation phase. That the formation

of a stable nucleus is not rate-limiting accounts for the lack of protein concentration-dependence on the lag phase of polymerization, as the dissociation rate would be independent of protein concentration. Such a model also explains the protein concentration-dependence on the rate of fibril formation but the lack of effect on the rate and lag phase of fibril formation of adding preformed seeds to the reaction mix.

Thus, the precursor to fibril formation is the dissociated species rather than the oligomeric form of the protein. The reasoning for this is two-fold. Firstly, seeding of the reaction with up to 20% w/w preformed fibrils did not have an effect on the rate of fibril formation. If the oligomer were the precursor to fibril formation than it is foreseeable that seeding at low concentrations (i.e. 0.5 – 2% w/w) would not have a significant effect on the lag phase or rate of fibril formation since the number of seeds would be relatively small compared to the number of pre-existing nuclei (in the form of oligomers). However, at higher concentrations the number of seeds would be relatively high and so one would expect to see an observable change in the lag phase and rate of fibril elongation since the number of amyloidogenic oligomers would not be limiting. Secondly, it is more likely that the dissociated species would enable the ordered aggregation required for fibril formation to proceed since it would expose the hydrophobic 'legs' that we have identified make up the protease-resistant core of the fibrils (see below). In contrast, in the oligomeric form of the protein the hydrophobic regions are thought to be internally buried (23,40), which would limit their ability to associate to form fibrils.

In the majority of studies describing fibril formation by a specific precursor protein, aggregation has been shown to involve solution conditions that disrupt the native structure and therefore facilitate unfolding. Thus, a critical component of the general model for amyloid fibril formation is the requirement for the protein to unfold, at least partially, as a precursor to amyloid nucleation (1) (see Fig. 7). However, for many systems involved with disease (e.g. Alzheimer's, Parkinson's), the peptides and proteins that have been found to aggregate adopt essentially unstructured conformations in their normal biological state, i.e. they are 'natively unfolded' (10). Thus, aggregation does not need to be preceded by the unfolding of a globular protein, but rather, may proceed by direct self-

assembly of the particular protein or peptide from an ensemble of unstructured conformations (10). Moreover, whilst dissociation of a protein from an oligomeric form can generate an amyloid precursor due to the change in solution conditions (e.g. for acid denaturation of the transthyretin tetramer to form monomers (41)), the possibility that amyloid fibril formation primarily occurs due to a 'native' amyloidogenic species dissociating from its binding partner or parent oligomer, under physiological conditions, has not been fully explored.

In the case of κ -casein, it is reasonable to speculate that the generation of the amyloidogenic intermediate may simply involve dissociation from its oligomeric state, i.e. quaternary structure destabilization. In the absence of the other casein proteins, monomeric forms of κ -casein are in dynamic equilibrium with a large oligomeric species (22). Treatments known to increase the rate of dissociation, such as heat and reduction, increase the rate of amyloid fibril assembly (13). It remains to be determined the extent of unfolding undergone by these dissociated forms in the generation of an amyloidogenic precursor. However, it is likely to be limited, since modelling (24) suggests that the dissociated κ -casein monomer samples a structural conformation that is an ideal β -sheet containing precursor for fibril formation, negating the need for further unfolding.

The rate-limiting species formed during amyloid fibril formation by RCM κ -CN is highly amyloidogenic - As part of our proposed model for the mechanism by which RCM κ -CN forms fibrils, the dissociated species is highly amyloidogenic, and so is readily incorporated into nuclei or growing fibrils when it becomes available. Thus, there is no free pool of this amyloidogenic species available to be incorporated into exogenous seeds as occurs in the simple nucleation-dependent model. We therefore sought to identify the regions of the protein incorporated into the protease-resistant core of fibrils formed by RCM κ -CN and investigated whether these help to explain its amyloidogenic nature. Our findings indicate that the region of RCM κ -CN from Tyr²⁵ to Lys⁸⁶ represents the protease-resistant β -sheet core of amyloid fibrils formed by κ -casein. This region correlates well with that predicted to fold into a well-defined conformation by protein disorder algorithms and is aggregation prone (see Fig. 6). Whilst κ -casein is regarded as a natively unstructured protein, since its monomer adopts

native molten-globule-like states, i.e. it has ill-defined tertiary structure whilst retaining some secondary structural elements (19), theoretical 3-D modelling of the protein has suggested that it may sample a "horse and rider" conformation (24). In this model, the region from Lys²¹ to Phe⁵⁵ is proposed to adopt an anti-parallel β -sheet formation in which the β -sheets are separated by proline turns (24). These 'legs', due to their high degree of hydrophobicity and exposure, would make an ideal site for sheet-sheet interactions. Under physiological conditions such interactions would presumably occur with hydrophobic domains of other caseins, facilitating its incorporation into oligomeric species such as milk micelles. However, this region of the protein would also favor β -sheet stacking with other κ -casein subunits, and therefore facilitate amyloid fibril assembly.

The assembly of fibrils formed from RCM κ -CN - Our studies show that a significant portion of the full-length protein is protected from proteolytic cleavage once it has adopted the fibrillar conformation, even when a relatively non-specific protease such as proteinase K is used for the digestion. This is similar to what has been reported for other amyloid forming systems, e.g. up to 89 residues (residues 10-99) of β 2-microglobulin were found to be resistant to degradation by pepsin (42). It is highly likely that a single κ -casein monomer may contribute more than one β -strand to the fibril structure, in a similar manner to that shown for the amyloid beta peptide (A β 1-40) (43) and the C-terminal fragment of the HET-s protein (44). Moreover, as the protease-resistant region of fibrillar κ -casein is continuous, the loops joining these strands are unlikely to be solvent exposed. Further structural studies of fibrils formed by κ -casein are required to confirm these proposals. It was unexpected that treatment with trypsin (a relatively specific protease) was able to cleave closer to the core of the amyloid fibrils than proteinase K (a relatively non-specific protease). This may reflect differences in the residue accessibility of the two proteases (due to packing of the fibrils) due to variations in their size and mode of action.

Upon fibril assembly, κ -casein undergoes structural rearrangement within the protein, particularly to the region predicted to have a high propensity to form β -sheet (24). Indeed, our spectroscopic data indicate that,

whilst some regions of the protein are not involved in fibril formation (as assessed by NMR spectroscopy – H Ecroyd, DM Williams and JA Carver, unpublished results), others undergo major rearrangements, leading to a significant increase in the β -sheet content of the protein (as shown by FTIR spectroscopy, see Fig. 2 and Table 1). Moreover, whilst we suggest that the inherent propensity for κ -casein to form amyloid fibrils originates from regions of the monomeric state that facilitate β -sheet stacking, the differences in the component bands of the amide I region of the FTIR spectra attributable to β -sheet suggest that the β -sheet domains in the native and fibrillar forms are structurally dissimilar. Thus, amyloid fibril growth does not simply involve the direct stacking of native β -sheets from individual precursors, but instead is most likely preceded by a reorganization of these regions, as has been suggested for other β -sheet containing proteins (45).

Conclusions - Here we show that fibril formation by RCM κ -CN is not governed by the simple nucleation-dependent model that has been proposed for most other amyloid fibril-forming systems studied to date (Fig. 7). Instead, we propose that the rate of fibril formation by RCM κ -CN is limited by the dissociation of an amyloidogenic precursor from an oligomeric

state (Fig. 7). In milk, the amount of this amyloidogenic species would be very low because of the interaction of κ -casein with itself (e.g. via intermolecular disulfide bonding) and with the other caseins in milk (e.g. in milk micelles). Nonetheless, the proteinaceous deposits found within *corpora amylacea* of mammary tissue have been shown to be amyloid in nature (14-16) and peptides corresponding to casein fragments have been isolated from within these amyloid-like deposits (46,47). Moreover, intracellular fibrillar bundles have also been identified in epithelial cells that surround the deposits of *corpora amylacea* (17,48). These are sites at which casein synthesis and secretion is high and the reducing environment of the cell would favor reduced forms of these proteins. Thus, it is possible that the dissociation of amyloidogenic precursors of κ -casein from milk micelles does occur under some circumstances, resulting in fibrillar deposits within these tissues. That amyloidoses associated with κ -casein fibril formation are not more prevalent *in vivo* is probably a testament to the interaction between the casein proteins, which prevent the release of this κ -casein amyloidogenic precursor and thereby inhibit large-scale κ -casein fibril formation through their chaperone-like activity (13,49).

References

1. Dobson, C. M. (2004) *Methods* **34**, 4-14
2. Chiti, F., and Dobson, C. M. (2006) *Annu Rev Biochem* **75**, 333-366
3. Harper, J. D., and Lansbury, P. T., Jr. (1997) *Annu Rev Biochem* **66**, 385-407
4. Stefani, M., and Dobson, C. M. (2003) *J Mol Med* **81**, 678-699
5. Guijarro, J. I., Sunde, M., Jones, J. A., Campbell, I. D., and Dobson, C. M. (1998) *Proc Natl Acad Sci U S A* **95**, 4224-4228
6. Chiti, F., Webster, P., Taddei, N., Clark, A., Stefani, M., Ramponi, G., and Dobson, C. M. (1999) *Proc Natl Acad Sci U S A* **96**, 3590-3594
7. Fandrich, M., and Dobson, C. M. (2002) *Embo J* **21**, 5682-5690
8. Fandrich, M., Fletcher, M. A., and Dobson, C. M. (2001) *Nature* **410**, 165-166
9. Dobson, C. M. (1999) *Trends Biochem Sci* **24**, 329-332
10. Chiti, F., Stefani, M., Taddei, N., Ramponi, G., and Dobson, C. M. (2003) *Nature* **424**, 805-808
11. DuBay, K. F., Pawar, A. P., Chiti, F., Zurdo, J., Dobson, C. M., and Vendruscolo, M. (2004) *J Mol Biol* **341**, 1317-1326
12. Farrell, H. M., Jr., Cooke, P. H., Wickham, E. D., Piotrowski, E. G., and Hoagland, P. D. (2003) *J Protein Chem* **22**, 259-273
13. Thorn, D. C., Meehan, S., Sunde, M., Rekas, A., Gras, S. L., MacPhee, C. E., Dobson, C. M., Wilson, M. R., and Carver, J. A. (2005) *Biochemistry* **44**, 17027-17036
14. Reid, I. M. (1972) *J Comp Pathol* **82**, 409-413
15. Beems, R. B., Gruys, E., and Spit, B. J. (1978) *Vet Pathol* **15**, 347-352

16. Taniyama, H., Kitamura, A., Kagawa, Y., Hirayama, K., Yoshino, T., and Kamiya, S. (2000) *Vet Pathol* **37**, 104-107
17. Nickerson, S. C. (1987) *Cytobios* **51**, 81-92
18. Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., and Obradovic, Z. (2002) *Biochemistry* **41**, 6573-6582
19. Syme, C. D., Blanch, E. W., Holt, C., Jakes, R., Goedert, M., Hecht, L., and Barron, L. D. (2002) *Eur J Biochem* **269**, 148-156
20. Fox, P. F. (2003) Milk proteins: general and historical aspects. In: Fox, P. F., and McSweeney, P. (eds). *Advanced Dairy Chemistry*, Springer, New York
21. Groves, M. L., Wickham, E. D., and Farrell, H. M., Jr. (1998) *J Protein Chem* **17**, 73-84
22. Rasmussen, L. K., Hojrup, P., and Petersen, T. E. (1992) *Eur J Biochem* **207**, 215-222
23. Vreeman, H. J., Brinkhuis, J. A., and van der Spek, C. A. (1981) *Biophys Chem* **14**, 185-193
24. Kumosinski, T. F., Brown, E. M., and Farrell, H. M., Jr. (1993) *J Dairy Sci* **76**, 2507-2520
25. Cappai, R., Leck, S. L., Tew, D. J., Williamson, N. A., Smith, D. P., Galatis, D., Sharples, R. A., Curtain, C. C., Ali, F. E., Cherny, R. A., Culvenor, J. G., Bottomley, S. P., Masters, C. L., Barnham, K. J., and Hill, A. F. (2005) *Faseb J* **19**, 1377-1379
26. Fox, P. F., and McSweeney, P. L. H. (1998) *Dairy Chemistry and Biochemistry*. Blackie Academic and Professional, London, U.K.
27. Ecroyd, H., Meehan, S., Horwitz, J., Aquilina, J. A., Benesch, J. L., Robinson, C. V., Macphie, C. E., and Carver, J. A. (2007) *Biochem J* **401**, 129-141
28. Byler, D. M., and Susi, H. (1986) *Biopolymers* **25**, 469-487
29. Laemmli, U. (1970) *Nature* **227**, 680-685
30. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) *Anal Chem* **68**, 850-858
31. Nilsson, M. R. (2004) *Methods* **34**, 151-160
32. Susi, H., and Byler, D. M. (1986) *Methods Enzymol* **130**, 290-311
33. Li, X., Romero, P., Rani, M., Dunker, A. K., and Obradovic, Z. (1999) *Genome Inform Ser Workshop Genome Inform* **10**, 30-40
34. Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., and Dunker, A. K. (2001) *Proteins* **42**, 38-48
35. Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E. H., Man, O., Beckmann, J. S., Silman, I., and Sussman, J. L. (2005) *Bioinformatics* **21**, 3435-3438
36. Uversky, V. N., Gillespie, J. R., and Fink, A. L. (2000) *Proteins* **41**, 415-427
37. Fernandez-Escamilla, A. M., Rousseau, F., Schymkowitz, J., and Serrano, L. (2004) *Nat Biotechnol* **22**, 1302-1306
38. Pawar, A. P., Dubay, K. F., Zurdo, J., Chiti, F., Vendruscolo, M., and Dobson, C. M. (2005) *J Mol Biol* **350**, 379-392
39. Lopez de la Paz, M., and Serrano, L. (2004) *Proc Natl Acad Sci U S A* **101**, 87-92
40. de Kruif, C. G., and May, R. P. (1991) *Eur J Biochem* **200**, 431-436
41. Lai, Z., Colon, W., and Kelly, J. W. (1996) *Biochemistry* **35**, 6470-6482
42. Myers, S. L., Thomson, N. H., Radford, S. E., and Ashcroft, A. E. (2006) *Rapid Commun Mass Spectrom* **20**, 1628-1636
43. Petkova, A. T., Ishii, Y., Balbach, J. J., Antzutkin, O. N., Leapman, R. D., Delaglio, F., and Tycko, R. (2002) *Proc Natl Acad Sci U S A* **99**, 16742-16747
44. Ritter, C., Maddelein, M. L., Siemer, A. B., Luhrs, T., Ernst, M., Meier, B. H., Saupe, S. J., and Riek, R. (2005) *Nature* **435**, 844-848
45. Zandomenighi, G., Krebs, M. R., McCammon, M. G., and Fandrich, M. (2004) *Protein Sci* **13**, 3314-3321
46. Niewold, T. A., Murphy, C. L., Hulskamp-Koch, C. A., Tooten, P. C., and Gruys, E. (1999) *Amyloid* **6**, 244-249
47. Claudon, C., Francin, M., Marchal, E., Straczek, J., Laurent, F., and Nabet, P. (1998) *Tissue Cell* **30**, 589-595
48. Nickerson, S. C., Sordillo, L. M., Boddie, N. T., and Saxton, A. M. (1985) *J Dairy Sci* **68**, 709-717
49. Morgan, P. E., Treweek, T. M., Lindner, R. A., Price, W. E., and Carver, J. A. (2005) *J Agric Food Chem* **53**, 2670-2683

Footnotes

* We thank Dr. Lyn Waterhouse (Medical School, University of Adelaide) for assistance with TEM. We also thank Dr Michele Vendruscolo (Department of Chemistry, University of Cambridge) for the data pertaining to the aggregation propensity of κ -casein using the modified version of the Zyggregator algorithm and Dr Sarah Meehan (Department of Chemistry, University of Cambridge) for helpful discussions about some aspects of this work.

This work was supported by grants from the National Health and Medical Research Council (NHMRC) of Australia, the Australian Research Council (JAC and PH) and Postgraduate scholarships from Dairy Australia (DCT), an Australian Postgraduate award (DMW) and the University of Adelaide (TK). HE is supported by a NHMRC Peter Doherty Fellowship and GD is supported by a NHMRC CJ Martin Fellowship.

The abbreviations used are: RCM κ -CN, reduced and carboxymethylated κ -casein; TEM, transmission electron microscopy; ThT, thioflavin T.

Figure legends

Fig. 1. Fibril formation of RCM κ -CN (0 - 10 mg/mL) as monitored by (A and B) ThT fluorescence and (D) light scattering at 340 nm. Samples were incubated in 50 mM phosphate buffer, pH 7.2 at 37°C and monitored over time. (C) A plot of the $\ln[\text{RCM}\kappa\text{-CN}]$ against $\ln[\text{reaction rate}]$. The reaction rate was determined from the initial rate of increase in ThT fluorescence from data such as shown in (A and B). In (D), the initial rate of the increase in light scattering has been extrapolated to determine the lag phase of the reaction (dotted lines). (E) The effect on ThT fluorescence of decreasing the temperature to 32°C on fibril formation of RCM κ -CN (0 - 10 mg/mL). (F) The effect of seeding on the rate of fibril formation by RCM κ -CN. A sample of RCM κ -CN (2.5 mg/mL) was incubated in 50 mM phosphate buffer, pH 7.2 at 37°C with or without the addition of preformed fibrils (seeds) at a 0.5% or 5.0% w/w ratio (Note – the seeded curves overlay each other in the graph). Inset – the addition of seeds to a sample of α -synucleinA53T incubated at 2 mg/mL in 50 mM phosphate buffer, 100 mM NaCl, pH 7.4 at 37°C with shaking (● – no seed added, ○ - 0.5% w/w seed added, ▼ – 2.0% w/w seed added).

Fig. 2. Analysis of RCM κ -CN before ($t = 0$ min) and after ($t = 1000$ min) incubation under conditions which promote amyloid formation fibril by (A) CD and (B) FTIR spectroscopy. Deconvolved FTIR data showing the individual component bands of the spectra for (C) RCM κ -CN before ($t = 0$ min) and (D) RCM κ -CN after ($t = 1000$ min) incubation.

Fig. 3. RCM κ -CN fibrils have a protease-resistant core. (A) SDS-PAGE analysis of native (N) and fibrillar (F) RCM κ -CN following treatment with trypsin for 0, 0.5, 1, 2 and 4 hr. Transmission electron micrographs of RCM κ -CN before (B) and after (C) incubation with trypsin for 4 hr. Scale bar represents 500 nm.

Fig. 4. Digestion of RCM κ -CN fibrils with proteinase K. (A) SDS-PAGE analysis of native (N) and fibrillar (F) RCM κ -CN following treatment with proteinase K for 0, 0.5, 1 and 2 hr. The 10 kDa and 8 kDa bands identified by mass spectrometry and shown in Table 1 are denoted by (*). Transmission electron micrographs of RCM κ -CN before (B) and after (C) incubation with proteinase K for 4 hr. Scale bar represents 200 nm.

Fig. 5. Identification of the protease-resistant core of amyloid fibrils formed by RCM κ -CN using mass spectrometry. (A) MALDI-TOF MS spectrum of peptide fragments present following treatment of RCM κ -CN fibrils with trypsin and their subsequent disaggregation and elution with 60% acetonitrile using ZipTip clean-up. (B) MALDI-TOF spectrum of the tryptic peptides from the in-gel

digest obtained from the Coomassie blue stained SDS-PAGE band that was identified following treatment of RCM κ -CN fibrils with trypsin.

Fig. 6. A summary of the key regions of κ -casein with regard to its amyloid fibril-forming ability. The diagram indicates the regions of the protein predicted to form β -sheet by 3D-modelling (24), the protease-resistant fibrillar core fragments generated by digestion with proteinase K and trypsin and identified by mass spectrometry, the region of the protein predicted to be ordered by the algorithms PONDR (33,34) and FoldIndex (35,36), the most amyloidogenic regions of the protein as predicted by TANGO (37) and a modified version of Zyggregator (38) (Tartaglia *et al.*, in preparation), and the region of the protein in which amino acid sequences match to those considered amyloidogenic as proposed by Lopez de la Paz and Serrano (39). In the latter, the (*) indicates that this amino acid sequence is followed by a proline residue which is considered an amyloid fibril breaker.

Fig. 7. A proposed model to describe differences in the process of fibril formation by proteins that follow a simple nucleation-dependent mechanism compared to fibril formation by κ -casein. In the nucleation-dependent mechanism, fibril formation commences with the unfolding of a native protein, forming a pool of partially folded intermediates, a process that is reversible. The partially folded intermediates are able to reversibly associate with each other until they reach a critical size/mass at which a stable nucleus is formed. The formation of this nucleus from the partially folded intermediates is slow and rate-limiting in the overall process of fibril formation. The time required for the formation of this nucleus is referred to as the lag phase which is protein concentration dependent since increasing the concentration increases the number of intermediates that can associate to form a nucleus. Fibril elongation then proceeds via the addition of intermediates to the growing nucleus. The mechanism also explains how seeding the reaction increases the reaction rate and decreases the lag phase since addition of preformed fibrils overcomes the time required to form nuclei (orange arrow) and the concentration of partially folded intermediates is not limiting. In contrast, fibril formation by κ -casein depends on the dissociation of a highly amyloidogenic species from its oligomeric state and this is the rate-limiting step in the fibril-forming pathway. Once released, these amyloidogenic species rapidly associate to form stable nuclei and, since the dissociation constant is independent of protein concentration, the lag phase of fibril formation by κ -casein is independent of the protein concentration. Fibril formation proceeds through a similar mechanism of β -sheet stacking as with the nucleation-dependent model, however, addition of preformed seeds to the reaction does not increase the reaction rate as the supply of the amyloidogenic species for fibril elongation is also limited by the dissociation constant (orange arrow).

Table 1: Estimations of secondary structural elements of RCMκ-CN before and after amyloid fibril formation as determined by FTIR analysis.

| | Before | | After | |
|--------------------|--------------------|-----------------|--------------------|-----------------|
| | Peak (cm-1) | Area (%) | Peak (cm-1) | Area (%) |
| Side Chain | 1612 | 10 | 1613 | 11 |
| β-structure | 1628 | 17 | 1628 | 22 |
| | | | 1639 | 9 |
| | | | 1677 | 10 |
| Disordered | 1648 | 48 | 1653 | 46 |
| Turns/Bends | 1670 | 20 | 1688 | 2 |
| | 1685 | 4 | | |

Table 2: Identification of peptide fragments of RCM κ -CN by time of flight mass spectrometry.

| Peptide source | Matching tryptic peptide fragment | Expected average mass [M+H] ⁺ in Da | Observed average mass ^a [M+H] ⁺ in Da | Expected monoisotopic mass [M+H] ⁺ in Da | Observed monoisotopic mass ^b [M+H] ⁺ in Da | Δ mass in ppm |
|--|-----------------------------------|---|--|--|---|-------------------------|
| Trypsin treated | Y[25-86]K | 7208.33 | 7207.66 | | | 93 |
| disaggregated fibrils | Y[35-97]R | 7208.24 | | | | 80 |
| Trypsin treated fibrils - in-gel digest of 7 kDa SDS-PAGE band | Y[35-68]R | | | 4010.06 | 4009.90 ^c | 40 |
| | S[69-86]K | | | 1980.08 | 1980.03 ^c | 25 |
| | Y[25-34]R | | | 1251.70 | 1251.70 ^c | 0 |
| Proteinase K treated fibrils – in-gel digest of 10 kDa SDS-PAGE band | Y[35-68]R | | | 4010.06 | 4010.03 | 7 |
| | S[87-111]K | | | 2872.37 ^d | 2872.42 | 17 |
| | F[17-34]R | | | 2188.21 | 2188.22 | 5 |
| | S[69-86]K | | | 1980.08 | 1980.09 | 5 |
| | Y[25-34]R | | | 1251.70 | 1251.73 | 24 |
| Proteinase K treated fibrils – in-gel digest of 8 kDa SDS-PAGE band | Y[35-68]R | | | 4010.06 | 4010.03 | 7 |
| | F[17-34]R | | | 2188.21 | 2188.22 | 5 |
| | S[69-86]K | | | 1980.08 | 1980.09 | 5 |
| | Y[25-34]R | | | 1251.70 | 1251.73 | 24 |

^a - data obtained in linear mode represent average masses of the [M+H]⁺-ions in Da

^b - data obtained in reflector mode represent monoisotopic masses of the [M+H]⁺-ions in Da

^c – the sequence identity of these peptides was confirmed using tandem mass spectrometry

^d – includes oxidation of two methionines to sulphoxides with subsequent mass change of 32.0

Figure 1

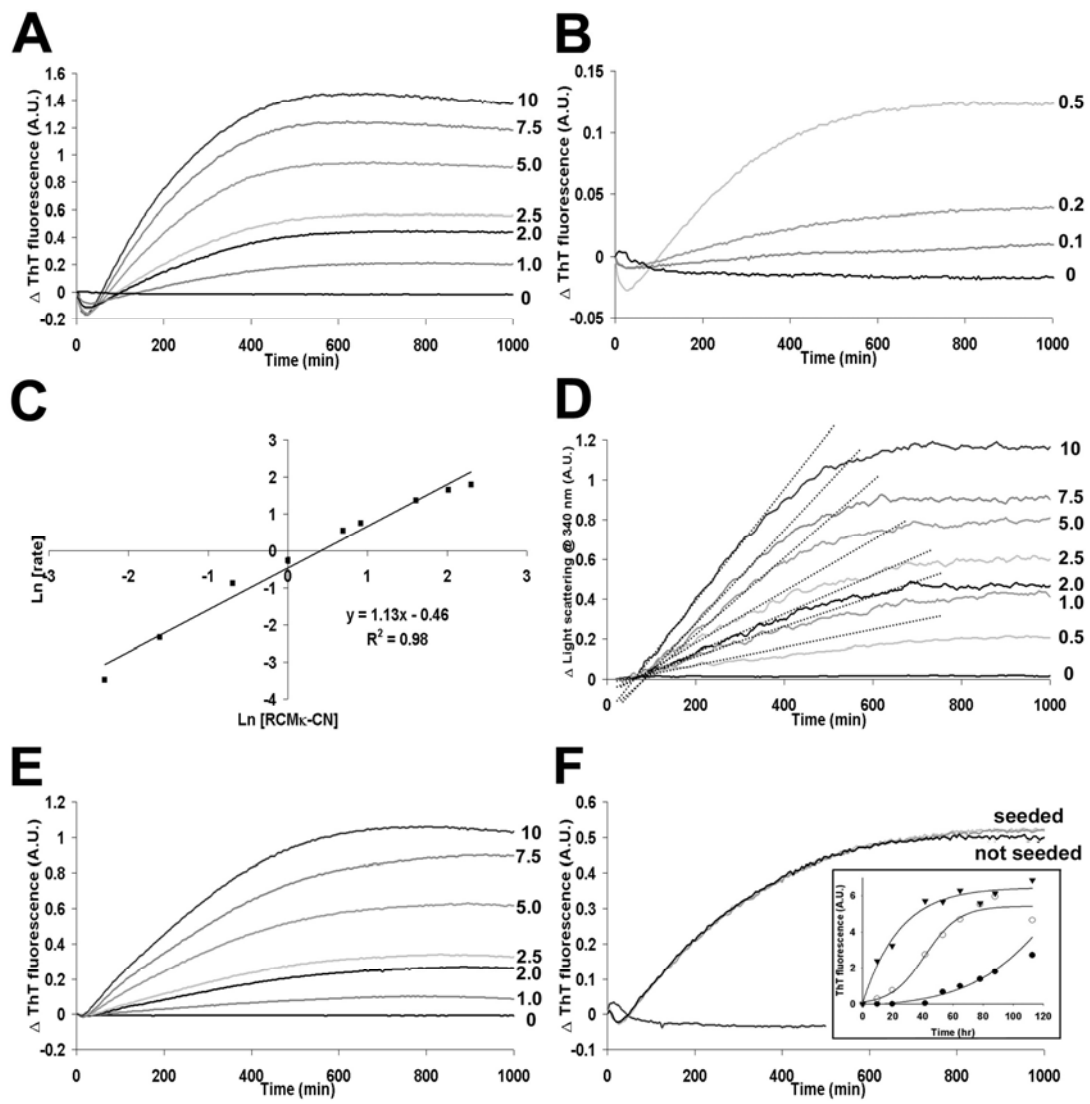


Figure 2

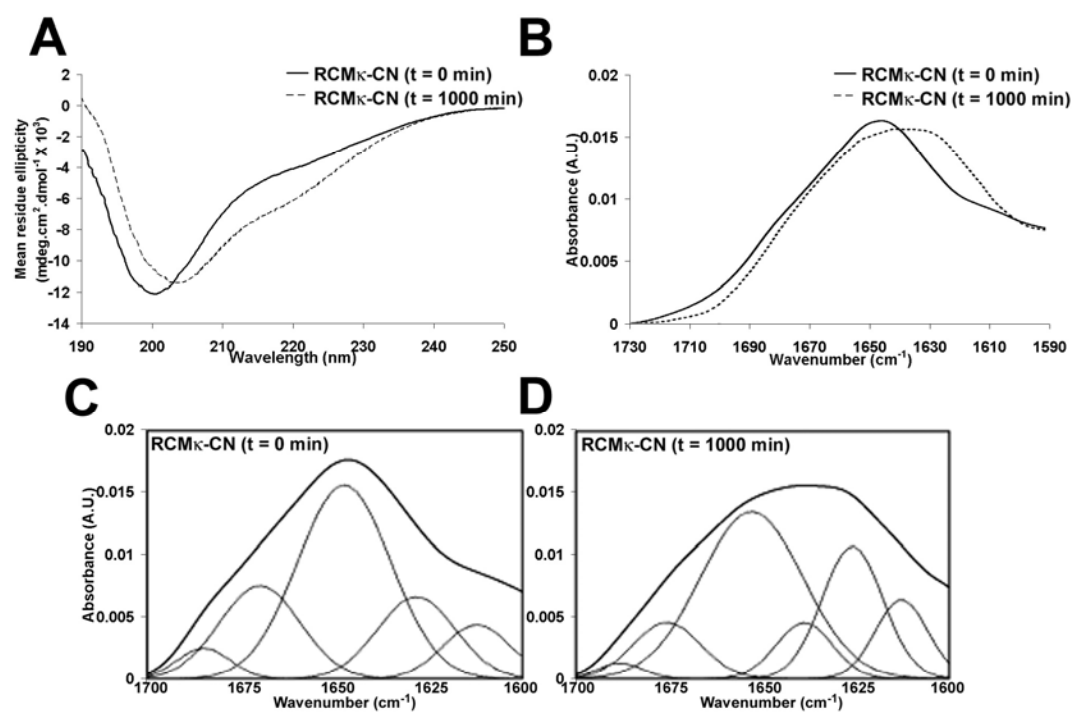


Figure 3

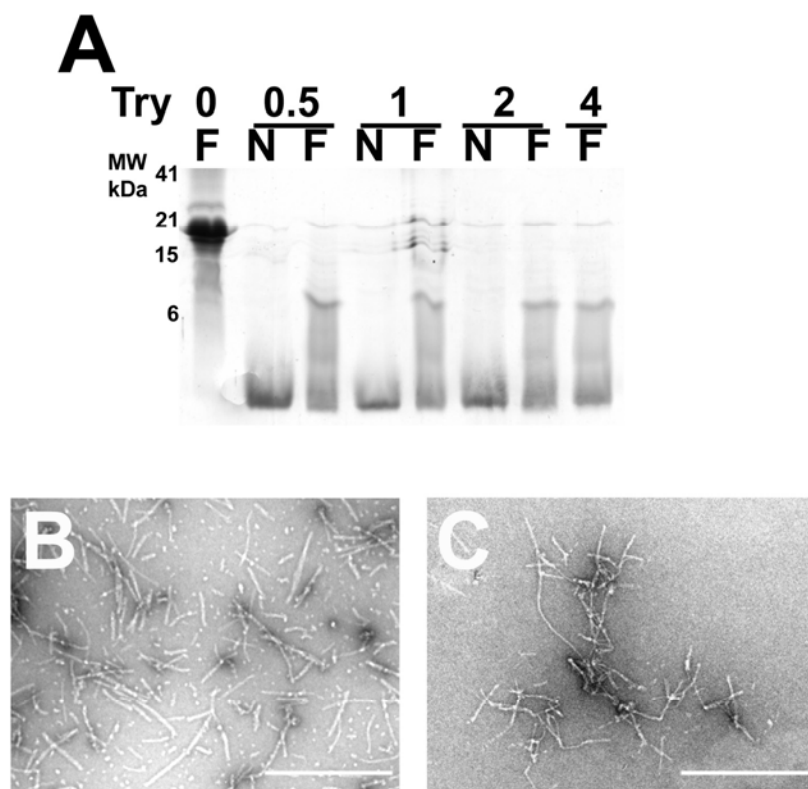


Figure 4

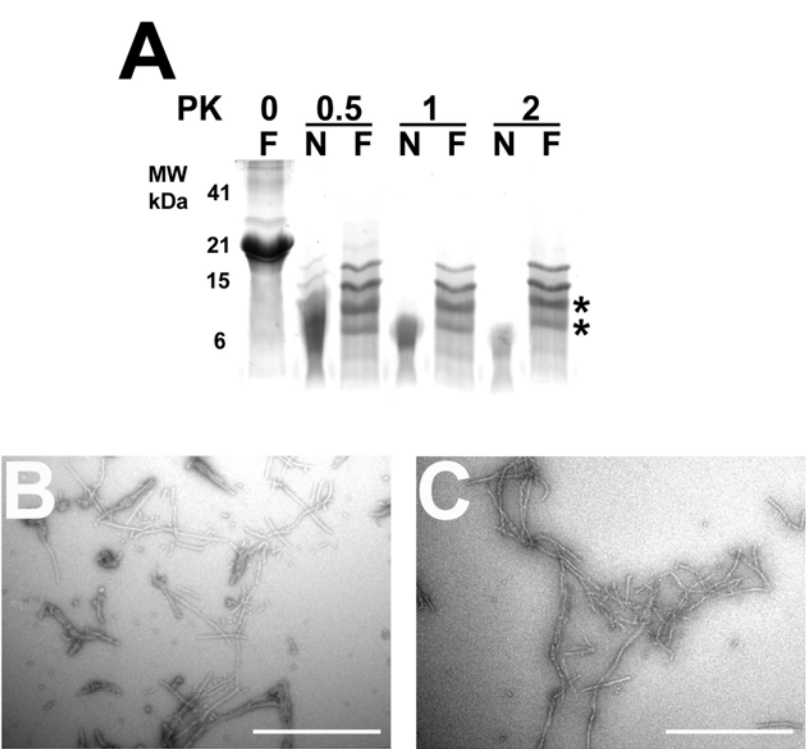


Figure 5

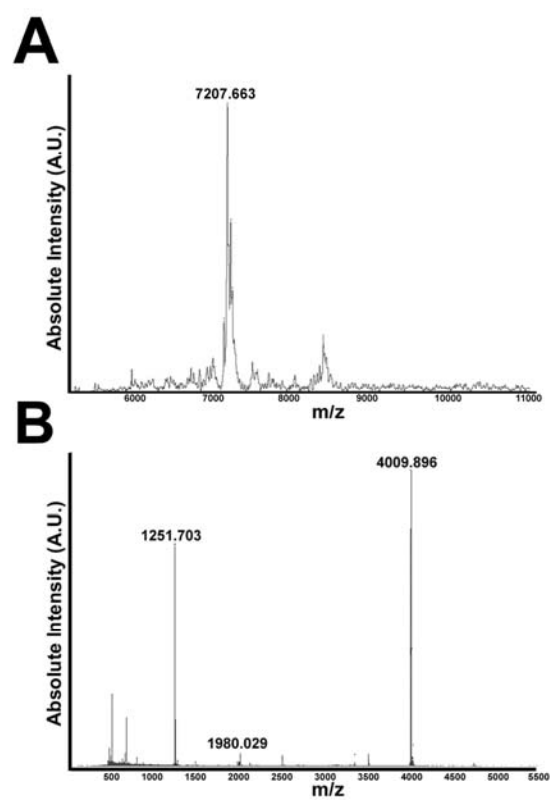


Figure 6

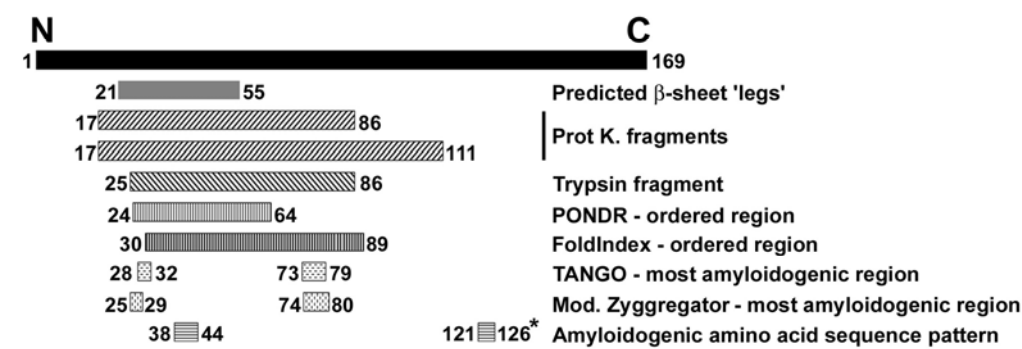
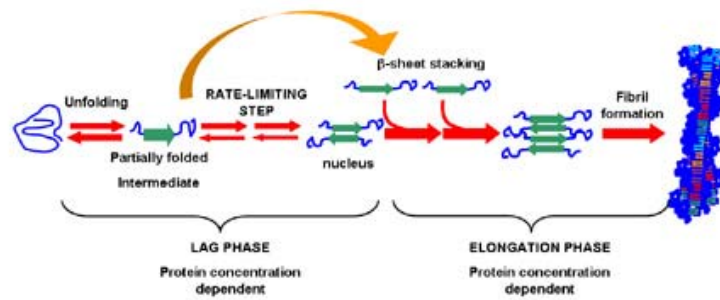


Figure 7

GENERIC NUCLEATION-DEPENDENT MODEL OF AMYLOID FIBRIL FORMATION



MODEL OF AMYLOID FIBRIL FORMATION BY κ -CASEIN

