



UNIVERSITY  
OF WOLLONGONG  
AUSTRALIA

University of Wollongong  
**Research Online**

---

Faculty of Science, Medicine and Health - Papers

Faculty of Science, Medicine and Health

---

2015

# Casein structures in the context of unfolded proteins

David C. Thorn

*University of Liege*

Heath Ecroyd

*University of Wollongong, [heathe@uow.edu.au](mailto:heathe@uow.edu.au)*

John A. Carver

*Australian National University, [john.carver@anu.edu.au](mailto:john.carver@anu.edu.au)*

Carl Holt

*University of Glasgow, [carl.holt@glasgow.ac.uk](mailto:carl.holt@glasgow.ac.uk)*

---

## Publication Details

Thorn, D. C., Ecroyd, H., Carver, J. A. & Holt, C. (2015). Casein structures in the context of unfolded proteins. *International Dairy Journal*, 46 2-11.

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library:  
[research-pubs@uow.edu.au](mailto:research-pubs@uow.edu.au)

---

# Casein structures in the context of unfolded proteins

## **Abstract**

Caseins were among the first proteins to be recognised as functional but unfolded. Many others are now known, providing better models of casein behaviour than either detergents or folded proteins. Caseins are members of a paralogous group of unfolded phosphoproteins, some of which share the ability to sequester amorphous calcium phosphate through phosphate centres. Non-covalent interactions of caseins can be through Pro- and Gln-rich sequences. Similar sequences in other unfolded proteins can also form open and highly hydrated structures such as gels, mucus and slimes. Many unfolded proteins, including  $\kappa$ - and  $\alpha$ S2-caseins, can form amyloid fibrils under physiological conditions. The sequence-specific interactions that lead to fibrils can be reduced or eliminated by low specificity interactions among a mixture of caseins to yield, instead, amorphous aggregates. The size of amorphous whole casein aggregates is limited by the C-terminal half of  $\kappa$ -casein whose sequence resembles that of a soluble mucin.

## **Disciplines**

Medicine and Health Sciences | Social and Behavioral Sciences

## **Publication Details**

Thorn, D. C., Ecroyd, H., Carver, J. A. & Holt, C. (2015). Casein structures in the context of unfolded proteins. *International Dairy Journal*, 46 2-11.

# Casein structures in the context of unfolded proteins

David C. Thorn<sup>a</sup>, Heath Ecroyd<sup>b</sup>, John A. Carver<sup>c</sup> and Carl Holt<sup>d\*</sup>

<sup>a</sup> Centre for Protein Engineering, University of Liège, 4000 Liège, Belgium.

<sup>b</sup> School of Biological Sciences and Illawarra Health and Medical Research Institute, University of Wollongong, Wollongong, NSW 2522, Australia.

<sup>c</sup> Research School of Chemistry, College of Physical and Mathematical Sciences, The Australian National University, Canberra, ACT 0200, Australia.

<sup>d</sup> Institute of Molecular, Cell and Systems Biology, School of Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK.

Corresponding author: Carl Holt, RB413A Level B4, Joseph Black Building, University of Glasgow, Glasgow G12 8QQ, Tel. +441292317615. Email [carl.holt@glasgow.ac.uk](mailto:carl.holt@glasgow.ac.uk)

## Abbreviations

PP-II: poly-*L*-proline-II

Caseins were among the first proteins to be recognised as functional but unfolded. Many others are now known, providing better models of casein behaviour than either detergents or folded proteins. Caseins are members of a paralogous group of unfolded phosphoproteins, some of which share the ability to sequester amorphous calcium phosphate through phosphate centres. Non-covalent interactions of caseins can be through Pro- and Gln-rich sequences. Similar sequences in other unfolded proteins can also form open and highly hydrated structures such as gels, mucus and slimes. . Many unfolded proteins, including  $\kappa$ - and  $\alpha_{S2}$ -caseins, can form amyloid fibrils under physiological conditions. The sequence-specific interactions that lead to fibrils can be reduced or eliminated by low specificity interactions among a mixture of caseins to yield, instead, amorphous aggregates. The size of amorphous whole casein aggregates is limited by the C-terminal half of  $\kappa$ -casein whose sequence resembles that of a soluble mucin.

## *1. Introduction*

In his autobiography in 1991, Arthur Kornberg made his famous remark that he had never met a dull enzyme (Kornberg, 1991), reflecting a life spent in uncovering the intricate relationships between structure and function among these largely globular proteins. Caseins were recognised to be unfolded as early as the 1950s (Halwer, 1954; Kresheck, 1965) when there were few other examples known to science. A senior figure in casein research later described casein as simply a denatured protein with only a nutritional function (McMeekin, 1970). This nutritional function of the unfolded conformation was thought to be the ease of digestion of casein by proteinases but this attracted little interest in the mainstream of protein science. In the last two decades the number of structure-function studies on unfolded proteins has grown exponentially from just a few per annum in the 1990s to thousands per annum today so that the functional roles of unfolded proteins in biology are now legion (Dyson & Wright, 2005; Rose, 2002; Tompa, 2012; Uversky & Dunker, 2010). In our view, casein research will benefit greatly from deep insights into the relationships between structure and function recently gained from studies on other unfolded proteins. In this review, we take some steps in this direction by considering the function of individual caseins and casein peptides in the sequestration of amorphous calcium phosphate to form thermodynamically stable nanoclusters. This ability is shared with osteopontin and probably with some other, closely related, unfolded phosphoproteins where the unfolded conformation is essential. The ability of individual caseins and casein peptides to form amyloid fibrils is shared with many other unfolded proteins and peptides. The ability of mixtures of different caseins to inhibit the growth of amyloid fibrils is an example of how promiscuous interactions of low sequence specificity can compete successfully with the highly specific interactions needed to form the fibrils.

In the past, the self-association of individual caseins and the association of casein mixtures in the absence of calcium ions or calcium phosphate have been likened to that of detergent molecules in which the hydrophobic effect (Cramer & Truhlar, 1992; Kauzmann, 1959) is assumed to provide the driving force (Horne, Lucey, & Choi, 2007; Mikheeva, Grinberg, Grinberg, Khokhlov, & de Kruif, 2003; Payens & Vreeman, 1982). Comparison with similar unfolded proteins suggests that an alternative driving force is more likely which involves main-chain-to-main-chain interactions of low sequence specificity rather than the side-chain

interactions of the hydrophobic effect. Such interactions lead naturally to open, extended and highly hydrated structures rather than the compact, anhydrous, domains of detergent micelles or the interior of globular proteins (Holt, Carver, Ecroyd, & Thorn, 2013).

In this review we will focus on two non-nutritional functions of the casein micelle which reduce the threat that lactation poses to the lifetime reproductive success of the mother: the control of ectopic calcification and the prevention of amyloidosis in the mammary gland. A third function is the need to form an easily digested gel in the stomach of the neonate. These three functions explain much of what we know about caseins and the structure of the casein micelle.

### **Origins and structures of casein genes**

To quote the evolutionary biologist Theodosius Dobzhansky, "Nothing in biology makes sense except in the light of evolution" (Dobzhansky, 1973), a statement that is as true for caseins and the casein micelle as it is for anything else in biology.

Calcium phosphate mineralised tissues appeared more than 500 million years (My) ago. Concomitantly, a group of SCPPs (secreted, calcium (phosphate)-binding phosphoproteins) evolved (Kawasaki & Weiss, 2003). Members of this group are involved in every aspect of biomineralisation. The SCPP genes can be divided into two groups depending on whether or not they contain long exons encoding sequences that are rich in Pro and Gln residues (P,Q-rich sequences). The P,Q-rich sequences are sticky and encourage protein-protein interactions with low sequence specificity, sometimes called promiscuous interactions (Hsu, et al. 2013; Kay, Williamson, & Sudol, 2000). Promiscuous interactions that involve only main-chain interactions are largely independent of sequence and are recognised to be important in, for example, the action of molecular chaperones and in enabling individual domains in a cell signalling network to respond to a variety of protein ligands (Macias, Wiesner, & Sudol, 2002).

Caseins all contain sticky, P,Q-rich, sequences which is one reason why they readily associate with themselves and each other. Casein genes are all descendants of the gene ODOM "Odontogenic ameloblast-associated protein" (Kawasaki, Lafont, & Sire, 2011),

highly expressed in dental tissue, through two separate lineages. The  $\kappa$ -casein genes are derived from ODAM via FDCSP “Follicular dendritic cell secreted peptide” which is active in the immune system and is highly expressed in adenoidal tissues. All other caseins, the so-called calcium-sensitive caseins, are derived from ODAM via the bone-associated protein SCPPPQ1 “SCPP P,Q-rich 1”. Calcium-sensitive caseins and peptides containing a phosphate centre can sequester amorphous calcium phosphate to form thermodynamically stable complexes (Little & Holt, 2004).

The other group of SCPPs, also known as SIBLINGS (small integrin-binding ligand, N-linked glycoproteins) (Fisher, Torchia, Fohr, Young, & Fedarko, 2001), are more acidic and do not contain sticky sequences. Like caseins they contain phosphate centre-type sequences and some have longer, highly phosphorylated, sequences. Among the earliest SCPPs to evolve was osteopontin (Kawasaki & Weiss, 2003). It is found in almost all species, tissues and biofluids and has at least 6 distinct functions (Mazzali, et al. 2002; Scatena, Liaw, & Giachelli, 2007). It is the most abundant non-collagenous protein in the extracellular matrix of bone and is thought, as the name implies, to form a bridge between the mineral and osteocytes. A naturally occurring mixture of osteopontin phosphopeptides will form a type of calcium phosphate nanocluster in which the core of amorphous calcium phosphate is more basic, four times larger and more highly hydrated than the core of calcium phosphate nanoclusters sequestered by casein phosphopeptides (Holt, Sorensen, & Clegg, 2009). Chameleon SCPPPQ1 contains a casein phosphate-centre-type sequence, SASSSEE (<http://www.uniprot.org/uniprot/E0YCE6>) but in the rat (<http://www.uniprot.org/uniprot/D6QY17>) and mouse (<http://www.uniprot.org/uniprot/B9UIU9>) sequences it is modified to SGGSSSEQ (Kawasaki, 2009; Moffatt, Smith, Sooknanan, St-Arnaud, & Nanci, 2006). Phosphate centre-type sequences have been identified in three other SCPPs and in a number of other secreted phosphoproteins (Holt, Sorensen, & Clegg, 2009). The ability of caseins to sequester amorphous calcium phosphate is therefore shared with, and derived from, other SCPPs.

The first casein evolved more than 300 My ago in some stem amniote before the great divergence into synapsids, the mammalian lineage, and sauropsids, leading to birds, dinosaurs, turtles and crocodiles. Lactation probably originated in mammal-like reptiles such

as cynodonts at least 50 My later (Lefèvre, Sharp, & Nicholas, 2010; Lemay, et al. 2009; Oftedal, 2012). We have argued that the current biological function of caseins in milk is an adaptation of an antecedent function of caseins in the control of some aspect of biomineralisation (Holt & Carver, 2012). The four recognised casein orthologues were all established before true mammals evolved (Lefèvre, Sharp, & Nicholas, 2010; Rijnkels, Kooiman, de Boer, & Pieper, 1997). In the eutherian lineage a second type of  $\alpha_{S2}$ -casein is found in some species and in the monotremes a second type of  $\beta$ -casein has evolved (Lefevre, Sharp, & Nicholas, 2009) so that there are six known casein gene products. All milks contain casein micelles. Notwithstanding this, there is considerable variation in the composition of caseins among extant mammals (Martin, Cebo, & Miranda, 2013). Only  $\kappa$ -casein is found in all species and individuals, along with at least two, and sometimes as many as four, other casein gene products. In eutherian species these other gene products are orthologues of the bovine  $\alpha_{S1}$ -,  $\alpha_{S2}$ - and  $\beta$ -caseins. Because the bovine orthologues can be precipitated by the addition of calcium ions at physiological pH, they are known as calcium-sensitive caseins, even though calcium sensitivity has not been demonstrated for all the members of this diverse group. The proportions of the different caseins can vary widely in milks from different species (Holt & Carver, 2012; Lefevre, Sharp, & Nicholas, 2009; Lefèvre, Sharp, & Nicholas, 2010; Martin, Ferranti, Leroux, & Addeo, 2003; Oftedal, 2012). It appears from this that while casein micelles are necessary and  $\kappa$ -casein is essential, the micelles can be made in many different ways, normally using at least two different calcium-sensitive-type caseins.

Orthologues of casein genes are highly variable in sequence and in the number of exons but they tend to be less variable in the lengths of the individual exons and in the amino acid composition of the encoded sequences. The bauplan of a group of related genes is, as the name implies, a simplified representation based upon conserved characters of a group that serve to distinguish members from an outgroup. The conserved characters can be chosen from all that is known about the structure, variability and function of the genes. Also, a bauplan does not necessarily correspond to the structure of any particular member of the group. Nor will any particular member of the group possess all the features chosen for the bauplan. Nevertheless, the bauplan is an abstraction that is useful because it captures some important common properties of the set. In Fig. 1 we have attempted to portray the bauplan of



the set of calcium-sensitive casein genes and show how it differs from the bauplan of the set of calcium-insensitive,  $\kappa$ -casein, genes.

The coding region of all casein genes begins and ends with an untranslated exon and the other coding exons generate at least three functional motifs in the translated protein. For the calcium-sensitive caseins (Fig. 1a), these are (i) a signal sequence encoded by a single short exon, (ii) a calcium phosphate-binding motif comprising a phosphate centre and flanking sequences, all encoded by short exons and (iii) a P,Q-rich sequence encoded by a longer exon. Some of the calcium-sensitive caseins also have an additional type of short exon encoding the C-terminus of the translated sequence. Any particular calcium-sensitive casein may contain more than one calcium phosphate-binding motif and more than one sticky, P,Q-rich sequence. Any given flanking sequence in a calcium phosphate binding motif may be encoded by more than one exon, or if the phosphate centre is at the N-terminus, none. For the calcium-insensitive caseins (Fig. 1b), the translated motifs are (i) a signal sequence encoded by a single short exon and (ii) a longer exon encoding a P,Q-rich sequence in its 5' half with some relatively well-conserved Tyr residues but in its 3' half encoding a Pro-, Ser- and Thr-rich (P,S,T-rich) sequence. The C-terminal half is more hydrophilic and acidic, particularly after modification by glycosylation and phosphorylation.

The amino acid composition encoded by signal sequence exons, when represented by a polar plot (Figure 2a), has a strongly developed vertical axis of hydrophobic residues. The amino acid composition of the calcium phosphate-binding motif, by contrast, is rich in charged and hydrophilic residues giving it a strongly developed horizontal axis of Glu and Ser residues in a polar plot (Figure 2b). The polar plot of the third motif, encoded by the longer type of exon, is a sticky, P,Q-rich sequence. The polar plot of these sticky sequences also has a strongly developed horizontal axis (Figure 2c) but in this case the neutral Pro and Gln residues are the most abundant. This type of sequence is not hydrophobic as a comparison of Figures 2a and 2c demonstrates. Indeed, the reputation of caseins as hydrophobic proteins is completely undeserved (Holt, Carver, Ecroyd, & Thorn, 2013).

The long exon of  $\kappa$ -casein genes encodes a fairly typical P,Q-rich, sticky, sequence in the N-terminal half (Figure 2d) and a number of Tyr residues may enhance the stickiness by side-

chain interactions, but the composition of the residues from the C-terminal half (Figure 2e) is very similar to that of a soluble mucin such as MUC-7 (Figure 2f) in which Pro, Ser and Thr residues are most abundant. Thus, the stickiness of the N-terminal half is reduced by the C-terminal half, and further reduced by glycosylation and phosphorylation in the C-terminal half. Typically, the boundary between the two halves is marked by a cleavage site for an aspartate proteinase such as chymosin or pepsin. Cleavage of  $\kappa$ -casein by chymosin at this site leads to gelation in the first stage of manufacture of many types of cheese from bovine, caprine and ovine milk.

## *2. Caseins are unfolded proteins*

It has been known for 60 years that caseins are unfolded but functional proteins (Halwer, 1954; Kresheck, 1965; McMeekin, 1952, 1970), but this knowledge had little impact on ideas of protein structure and function outside the casein field. A later proposal relating structure to function was that the unfolded conformation was advantageous in inhibiting calcium phosphate precipitation from solution so that it did not progress beyond the nucleation stage (Holt & Sawyer, 1993). The study of unfolded proteins has mushroomed of late to the point where it is now widely accepted that many such proteins have important biological functions. For example, it is now recognized that caseins are typical SCPPs in that they all have a largely flexible, unfolded, conformation (Fisher, Torchia, Fohr, Young, & Fedarko, 2001; Holt, Sorensen, & Clegg, 2009). We have expressed the view that unfolded proteins in general, and the other SCPPs in particular, provide better models for understanding the behaviour of caseins than either globular proteins or surfactants (Holt, Carver, Ecroyd, & Thorn, 2013). One consequence of applying this new knowledge is that the old idea of the importance of the hydrophobic interaction in casein chemistry needs to be revised.

A striking example of how caseins differ from globular proteins is that in the former, disulphide bridges are mostly intramolecular whereas in the latter, they are mostly intermolecular. Cys residues are normally highly conserved in globular proteins but this is not the case in caseins. Cys residues are confined to  $\kappa$  and  $\alpha_{S2}$  in the cow and to  $\kappa$ -casein alone in the rabbit. However, they are found in human  $\kappa$  and  $\alpha_{S1}$  and in all the main caseins of rat and mouse (Bouguyon, Beauvallet, Huet, & Chanat, 2006). The highest degree of conservation is arguably that of the Cys residue near the N terminus of the sequence encoded by exon 4 in  $\kappa$ -

caseins. This occurs in position 9, 10 or 11 of the mature eutherian and metatherian  $\kappa$ -caseins but is absent in the two monotreme  $\kappa$ -caseins (Lefevre, Sharp, & Nicholas, 2009). In other positions the Cys residues are conserved only among closely related species. In general, the Cys residues are in the form of cystine. For example, bovine  $\kappa$ -casein exists as a monomer with an intramolecular bridge between Cys-11 and Cys-88 but mostly as a range of homo-oligomers formed by intermolecular disulphide bridges of all three possible types (Holland, Deeth, & Alewood, 2008; Rasmussen, et al. 1999). The bovine  $\alpha_{S2}$ -casein homodimer involves two intermolecular disulphide bridges but the two peptide chains can be oriented in parallel or antiparallel directions so that again, all three possible disulphide bridges are found (Rasmussen, et al. 1999). Whereas the bovine caseins form disulphide bridges by homotypic interactions, this is not generally the case. Disulphide-linked hetero-oligomers are found in human, rat and mouse milks (Bouguyon, et al. 2006; Rasmussen, Due, & Petersen, 1995). The role of disulphide bridging in casein micelle structure is an enigma. Disulphide bridging may have a role in forming or maintaining the integrity of casein micelles but the nature and extent of the intermolecular bonding is quite variable among species.

Unfolded proteins, at least at physiological temperature, are not random coils. Instead they tend to adopt the extended poly-L-proline type 2 (PP-II) conformation (Cubellis, Caillez, Blundell, & Lovell, 2005; Shi, Chen, Liu, & Kallenbach, 2006). Any of the 20 or so amino acid residues can be found in a PP-II conformation but Pro is particularly prone to adopt this conformation because of the restricted range of its backbone conformational angles. In this conformation the backbone is fully exposed to water which forms a stabilising hydration shell around it. In this respect too, Pro is favoured because its carbonyl moiety is a particularly good H-bond acceptor and it holds on tenaciously to its hydration shell, even when placed in a hydrophobic environment. The conformation is further stabilised by dipolar interactions along the backbone between N-H and carbonyl moieties so that they form a 3-residues per turn helix (Maccallum, Poet, & Milner-White, 1995). Gln residues are also frequently found in the PP-II conformation which they may sometimes stabilise by forming a side-chain-to-backbone H-bond between residues  $i$  and  $i+1$  (Adzhubei, Sternberg, & Makarov, 2013; Stapley & Creamer, 1999) but there are no main-chain-to-main-chain H-bonds in the PP-II structure.

Unfolded proteins are usually less hydrophobic and more highly charged than globular proteins so they do not readily self-associate (Uversky, Gillespie, & Fink, 2000). The P,Q-rich sequences of caseins readily adopt the PP-II conformation and do not have the high charge density of other parts of the sequence. Sequence complexity is low due to cryptic tandem repeats (Holt & Sawyer, 1993) so that many alternative but energetically similar interactions are possible. Because Pro-rich sequences are not readily desolvated, the PP-II conformation tends to be conserved and the resulting supramolecular structures are open, extended and highly hydrated gels, mucus, slimes and amorphous aggregates (Cubellis, Caillez, Blundell, & Lovell, 2005; Kay, Williamson, & Sudol, 2000; Shewry & Halford, 2002; Williamson, 1994). Pro-rich sequences in the PP-II conformation also feature prominently in the so-called promiscuous interactions of peptides binding to a shallow hydrophobic cleft in the SH3 and WW domains of signalling networks (Horita, et al. 1998; Macias, Wiesner, & Sudol, 2002; Siligardi, et al. 2012). Even in the hydrophobic cleft, the Pro-rich peptides retain interfacial water molecules (Martin-Garcia, Ruiz-Sanz, & Luque, 2012).

In the past, the self-association of caseins has been likened to that of detergent molecules in which the P,Q-rich sequences are supposed to be the equivalent of the hydrophobic tails. Structural studies on other unfolded proteins and peptides with Pro-rich sequences suggest that this analogy is a poor one because no compact and anhydrous domain results from the association process (Dalglish, 2011). The driving force may, in some cases (for example  $\beta$ -casein self-association), be largely entropic but the hydrophobic interaction of side chains and the formation of a main-chain-to-main-chain H-bonded network can have the same or a similar thermodynamic signature in terms of the changes in enthalpy, entropy, specific heat and molar volume (Cooper, 2000, 2005). Further development of this point will be made in section 6 when considering the role of promiscuous interactions in the action of caseins as molecular chaperones. Some further aspects to the argument were developed elsewhere (Holt, Carver, Ecroyd, & Thorn, 2013).

In summary, the sticky P,Q-rich sequences in caseins are not hydrophobic, they do not associate with each other by the hydrophobic effect as classically defined (Kauzmann, 1959) and they bear no similarity to the tails of detergent molecules. They do not condense into

dense dehydrated domains like the inside of globular proteins or detergent micelles but retain their solvation and form open, extended, highly solvated, gels, viscous solutions and amorphous aggregates. Some limited desolvation of side chains may be part of the predominantly main-chain-to-main-chain interactions as the demarcation of interactions into these two separate types is somewhat arbitrary (Mezei, Fleming, Srinivasan, & Rose, 2004; Syme, Dennis, Phillips, & Homans, 2007).

### ***3. Caseins can sequester amorphous calcium phosphate***

In the 1960s the presence of so-called colloidal calcium phosphate in casein micelles was already well known (McGann & Pyne, 1960) but its intimate relationship with the phosphate centre sequences of the calcium-sensitive caseins was not recognised until later (Aoki, Yamada, Tomita, Kako, & Imamura, 1987; Gagnaire, Pierre, Molle, & Léonil, 1996; Holt, Davies, & Law, 1986; Holt, et al. 1989; Ono, Ohtawa, & Takagi, 1994). The calcium phosphate-binding motif (Fig. 1) typically comprises a phosphate centre and either one (if near the N-terminus) or two flanking sequences of hydrophilic residues. A possible advantage of the unfolded conformation is that a high density of phosphate centres can be accommodated at the interface with the calcium phosphate. At first it was thought that adsorbed phosphopeptides provided a kinetic barrier to the further growth or aggregation of the sequestered nanoclusters (Holt & Sawyer, 1993; Holt, Wahlgren, & Drakenberg, 1996). However, further theoretical and experimental work demonstrated that an equilibrium complex was formed in which the radius of the calcium phosphate nanocluster depended on the strength of binding and density of packing of the phosphate centres at the interface (Clegg & Holt, 2009; Little & Holt, 2004). Evidence has since been presented that this phenomenon is not confined to milk or limited to caseins but may be of general physiological importance in the stabilisation of other biofluids and the control of biomineralisation in soft and hard tissues (Holt, Sorensen, & Clegg, 2009; Holt, 2013; Holt, Lenton, Nylander, & Teixeira, 2014).

### ***4. Caseins can form amyloid fibrils***

Caseins in general are predicted to be able to form amyloid fibrils (Holt & Carver, 2012) and bovine  $\alpha_{S2}$ - and  $\kappa$ -caseins will do so under physiological conditions (Farrell, Cooke, Wickham, Piotrowski, & Hoagland, 2003; Léonil, et al. 2008; Thorn, et al. 2005; Thorn,

Ecroyd, Sunde, Poon, & Carver, 2008). In this respect they are like many other unfolded proteins including their evolutionary ancestor ODAM (Murphy, et al. 2008). Another SCPP, ameloblastin, when expressed recombinantly in an unphosphorylated form, forms twisted, ribbon-like, structures with an average width and thickness of 18 and 0.34 nm (Wald, et al. 2013).

Arcs in the X-ray fibre diffraction pattern of an amyloid fibril bundle show that the characteristic  $\beta$ -sheet structure runs across the fibre axis at a shallow angle, probably forming a helical spiral. The detailed structure of a whole fibril is unsolved but a number of models have been proposed including the popular fibre bundle model of Fig. 5 in (Jiménez, et al. 2002) and the hollow fibre model in which a 20-residue Gln spiral is proposed for the amyloid form of the protein huntingtin, involved in Huntington's disease (Perutz, Finch, Berriman, & Lesk, 2002).

Higher resolution local structures have been proposed based on the crystal structures formed by amyloid-forming, so-called, steric zipper peptides (Nelson, et al. 2005; Sawaya, et al. 2007; Wiltzius, et al. 2008). In a steric zipper, two identical sequences form a tightly complementary, interdigitated, interface which, it is proposed, generates the spine of an amyloid fibril. It is the interdigitated side chains from the two complementary sequences that evoke the metaphor of a clothing zipper. Rosetta lattice energy calculations on the steric zipper structure formed by the hexapeptide NNQQNY have been used by Eisenberg's group to provide a fairly reliable method of testing whether different hexapeptides are able to form the zipper structure. The interaction energy calculations can be made on every possible six consecutive residues in a given protein to find those that are compatible with the steric zipper structure (Goldschmidt, Teng, Riek, & Eisenberg, 2010). Their finding is that almost all proteins contain at least one predicted steric zipper sequence. In globular proteins, steric zippers are usually confined to the interior and so can only react with each other when the protein is partially unfolded as a result of exposure to an elevated temperature or other source of stress. Unfolded proteins such as  $\alpha$ -synuclein, tau and huntingtin are among the most notorious amyloid-forming proteins, if only because of their association with Parkinson's, Alzheimer's and Huntington's neurodegenerative diseases, respectively. They cannot bury amyloidogenic sequences in their interior. Nevertheless, some unfolded proteins can act as

their own molecular chaperone so that amyloid fibrils only form after partial enzymatic digestion (Chiti & Dobson, 2006).

According to the Rosetta steric zipper predictions, all caseins that have been examined (Holt & Carver, 2012) contain amyloidogenic sequences. Some of the predicted zipper sequences are inconsistent with chemical modifications such as glycosylation, phosphorylation and disulphide bridging. With these exclusions, we find that steric zipper sequences with the highest propensity to form amyloid fibrils are located only in the sticky P,Q-rich sequences even though Pro itself cannot form part of the cross  $\beta$ -sheet structure. After allowing for secondary modification and the removal of the signal sequence, the predictions for the bovine caseins are that mature  $\beta$ -casein is the least likely to form amyloid but all the other bovine caseins are predicted to be able to form amyloid fibrils (Holt & Carver, 2012). Many of the predicted steric zipper sequences neither contain nor are near to a Cys residue so the occurrence of Cys residues in caseins, including the most highly conserved Cys residue in  $\kappa$ -caseins at positions 9, 10 or 11, cannot be a major factor in the prevention of amyloid fibril nucleation.

Experimentally,  $\kappa$ -casein readily forms amyloid fibrils under physiological conditions (Farrell, Cooke, Wickham, Piotrowski, & Hoagland, 2003; Léonil, et al. 2008; Thorn, et al. 2005), as does  $\alpha_{S2}$ -casein (Thorn, Ecroyd, Sunde, Poon, & Carver, 2008). Neither  $\beta$ - nor  $\alpha_{S1}$ -casein has been observed to form amyloid fibrils under physiological conditions.

We and others have proposed that the bovine  $\kappa$ -casein fibril core is located somewhere between Tyr25 and Lys86 (Ecroyd, et al. 2008; Ecroyd, Thorn, Liu, & Carver, 2010; Farrell, Cooke, Wickham, Piotrowski, & Hoagland, 2003) because, upon fibril formation, this region is resistant to proteolysis. Presumably this is because of the incorporation into, and thus burial of, at least part of this region in the  $\beta$ -sheet fibril core. Moreover, the segments Val48-Phe55 and Ala71-Val83 of bovine  $\kappa$ -casein each contain one or more overlapping steric zipper sequences (Holt & Carver, 2012). In  $\alpha_{S2}$ -casein, 2 zipper sequences, Asn83-Try89 and Gln94-Leu99, are present within the central P,Q-rich portion of the protein. Zipper sequences occur in the plasmin peptides Ala81-Val112, Ala81-Lys113, and Ala81-Arg125 from  $\alpha_{S2}$ -casein,

which have been isolated from mineralized, amyloid-like deposits (*corpora amylacea*) in bovine mammary tissue (Niewold, Murphy, Hulskamp-Koch, Tooten, & Gruys, 1999).

After digestion of casein micelles at constant pH with a broad-specificity proteinase, followed by high speed centrifugation, the pelleted fraction contains many phosphopeptides in a high molecular weight form because they are complexed to nanoclusters of sequestered amorphous calcium phosphate (Holt, Davies, & Law, 1986; Ono, Ohotawa, & Takagi, 1994). However, when the highly specific proteinase trypsin was used (Gagnaire, Pierre, Molle, & Léonil, 1996), a number of other, unphosphorylated, peptides were found either wholly or in part in the pelleted fraction. None of these peptides would have been pelleted to any significant extent unless they were in a higher molecular weight form. A detailed comparison of the pelleted peptide sequences showed that virtually all of them were from the P,Q-rich sequences and contained one or more steric zipper regions, see the Supplementary File, section S3.3 (<http://dx.doi.org/10.3168/jds.2013-6831>) of (Holt, Carver, Ecroyd, & Thorn, 2013). It is not yet known whether the pelleted, non-phosphorylated peptides are in a fibrillar or amorphous high molecular weight form.

Proteolytic digestion of amyloidogenic proteins can result in an increase in the rate of fibril formation. For example,  $\alpha$ -synuclein is nearly entirely in a PP-II conformation (Syme, et al. 2002) but has some significant long-range interactions between the negatively charged C-terminal region and the central amyloidogenic core (NAC) region (Bertoncini, et al. 2005). Structural perturbations that destabilize the interactions between these two portions of the protein molecule, including the deletion of the C-terminal region by selective proteolysis, appear to increase the exposure of the amyloidogenic NAC region and promote the formation of fibrils (Hoyer, Cherny, Subramaniam, & Jovin, 2004). Theoretical work and simulations show that a direct interaction of flanking, disordered regions with a central region prone to form aggregates is not necessary in order for the flanking regions to effect a reduction in the rate of fibril formation (Abeln & Frenkel, 2008; Hall, Hirota, & Dobson, 2005; Damien, Hall & Hirota, 2009). The observations of high molecular weight peptides in tryptic digests of casein micelles and of  $\alpha_{s2}$ -casein peptides in *corpora amylacea*, may be further examples of how the removal of flanking sequences in an unfolded protein enhances the aggregation of an amyloidogenic central sequence.



Experiments on the rate of fibril growth by purified bovine  $\kappa$ - or  $\alpha_{S2}$ -caseins have shown effects due to disulphide bridging. A mixture of disulphide-linked dimers and trimers of bovine  $\kappa$ -casein forms fibrils but not as readily as the fully reduced  $\kappa$ -casein (Thorn et al. 2005). So here, intermolecular disulphide bonding can be seen as having a protective effect, either directly or by stabilising non-fibrillar associated states of which the casein micelle can be seen as a natural example. With bovine  $\alpha_{S2}$ -casein, on the other hand, the disulphide-linked dimer is more amyloidogenic than the monomeric form (Thorn, Ecroyd, Sunde, Poon, & Carver, 2008). Unfortunately there are no data on amyloid fibril formation by caseins from other species.

The ready availability of caseins suggests that they, or peptides derived from them, could be used as commercial sources of amyloid fibrils for applications as bionanomaterials (Ecroyd, Garvey, Thorn, Gerrard, & Carver, 2013; Raynes & Gerrard, 2013).

### *5. Caseins can act as molecular chaperones*

An important mechanism responsible for ensuring that proteins attain and remain in their biologically active conformation is through the action of molecular chaperone proteins (Barral, Broadley, Schaffar, & Hartl, 2004; Hartl, Bracher, & Hayer-Hartl, 2011). Molecular chaperones such as the small heat-shock proteins (sHsps) and the extracellular chaperone clusterin typically bind to exposed residues in a partially folded state of a target protein to prevent it from forming an amorphous aggregate or amyloid fibril (Carver, Rekas, Thorn, & Wilson, 2003). All molecular chaperones can interact with a range of target proteins through protein-protein interactions that have low sequence specificity. Caseins, for example, can stabilize a range of unrelated globular proteins that have been partially unfolded by heat, chemical reduction, or UV-induced stress (Bhattacharyya & Das, 1999; Hassanisadi, et al. 2008; Koudelka, Hoffmann, & Carver, 2009; Matsudomi, Kanda, Yoshika, & Moriwaki, 2004; Morgan, Treweek, Lindner, Price, & Carver, 2005; Treweek, Thorn, Price, & Carver, 2011; Yong & Foegeding, 2010; Zhang, et al. 2005). Caseins can also limit or prevent the growth of amyloid fibrils. For example,  $\alpha_{S1}$ - and  $\beta$ -casein can prevent the formation of fibrillar structures by ovalbumin (Khodarahmi, Beyrami, & Soori, 2008) and the amyloid- $\beta$  peptide (Carrotta, et al. 2012).

The occurrence of refractory and potentially cytotoxic casein amyloid in the mammary gland would undoubtedly be dysfunctional. Fibril formation by  $\kappa$ -casein is cytotoxic to PC-12 cells grown in culture (Dehle et al. 2010). More recently, it has been shown that  $\alpha_{S2}$ -casein is also toxic to PC-12 cells and the human-derived neuron-like cell line SH-SY5Y [Abigail Regoeng (University of Adelaide, Australia), John A. Carver (Australian National University, Canberra) and Ian F. Musgrave (University of Adelaide), unpublished data]. So what is it that prevents caseins from forming amyloid? The answer to this question is that, in a mixture of different caseins, they stop each other through promiscuous interactions that lead to amorphous aggregates rather than amyloid fibrils. If the mixture includes  $\kappa$ -casein, the amorphous aggregates can be self-limiting in size.

Thus, the formation of fibrillar structures in solutions of either  $\alpha_{S2}$ - or  $\kappa$ -casein can be prevented by the addition of either  $\alpha_{S1}$ - or  $\beta$ -casein (Thorn, et al. 2005; Thorn, Ecroyd, Sunde, Poon, & Carver, 2008). It was thought at first that the tendency of  $\kappa$ - and  $\alpha_{S2}$ -caseins to form amyloid was countered by  $\beta$ - and  $\alpha_{S1}$ -caseins acting as molecular chaperones. However,  $\alpha_{S2}$ -casein also inhibits fibril formation by  $\kappa$ -casein, and vice versa (Treweek, Thorn, Price, & Carver, 2011). It appears therefore that each type of casein is able to inhibit, at least to some degree, the fibril-forming tendency of other caseins.

## ***6. Functional and dysfunctional processes in casein micelle assembly***

Harry Farrell wisely pointed out in a review article (Farrell, Qi, & Uversky, 2006) that a structure such as the casein micelle could only come about through functional interactions and the absence or limited influence of dysfunctional interactions.

### ***6.1 Control of dysfunctional ectopic calcification***

The first dysfunctional possibility is ectopic or pathological calcification of the mammary gland. This problem was solved by adapting methods already in use for 200 Myr before caseins evolved. Osteopontin phosphopeptides (and most likely phosphopeptides derived from other members of the group of SCPPs) can sequester amorphous calcium phosphate to form thermodynamically stable nanocluster complexes (Holt, Sorensen, & Clegg, 2009). Casein phosphopeptides form similarly stable complexes although the calcium phosphate

core differs in composition and size (Holt, Wahlgren, & Drakenberg, 1996; Holt, Timmins, Errington, & Leaver, 1998; Little & Holt, 2004). Solutions of the nanocluster complexes are undersaturated with respect to the first, amorphous, phase to precipitate from solution under physiological conditions and so the precipitation process cannot begin (Holt, 2013; Holt, Lenton, Nylander, & Teixeira, 2014). In practice, therefore, perfectly stable milk containing arbitrarily high concentrations of calcium and inorganic phosphate can be stored in the mammary gland for indefinitely long periods. In principle, stable milk can be made using sequestered amorphous calcium phosphate in the form of individual nanoclusters but instead, caseins contain the sticky P,Q-rich sequences that are partly responsible for the larger amorphous aggregates we call casein micelles.

## ***6.2 Control of dysfunctional amyloid formation***

A dysfunctional process could result from the formation of cytotoxic amyloid fibrils by one or more of the caseins. As discussed in section 5, in bovine milk, both  $\kappa$ -casein and  $\alpha_{S2}$ -casein are amyloidogenic but the highly ordered and cytotoxic fibrils do not normally form to any appreciable extent because the mixture of different caseins act on each other as molecular chaperones (Thorn, et al. 2005; Thorn, Ecroyd, Sunde, Poon, & Carver, 2008; Treweek, Thorn, Price, & Carver, 2011). Thus, the highly sequence-specific interactions needed to form the cross- $\beta$ -structure in the spine of an amyloid fibril can be reduced or eliminated by competition from a large number of competing interactions of the P,Q-rich casein sequences of low sequence specificity. The result is the highly hydrated amorphous casein micelle.

In some species, the number of casein gene products that combine to form a casein micelle is as high as five, whereas in others it is as low as three, and the proportions of each are highly variable (Baranyi, Brignon, Anglade, & Ribadeau-Dumas, 1995; Boumahrou, et al. 2009; Miranda, Mahe, Leroux, & Martin, 2004). For the first time in the history of casein research we have a possible explanation for why, in any particular type of milk, more than one calcium-sensitive casein is sometimes needed to form a casein micelle, i.e. to prevent amyloid fibrils from forming.

### ***6.3 Control of micelle size and gelation***

A third dysfunctional process could take place: the unlimited aggregation of the calcium-sensitive caseins to form a precipitate, especially at the high concentrations of casein present in the milk of most species. Casein micelle size can increase through two principal mechanisms. First, a number of the calcium phosphate nanoclusters can become linked together by bridging casein sequences having more than one phosphate centre. The evidence for this is the incomplete dissociation of casein micelles into individual nanoclusters by even high concentrations of urea (Holt, 1998). The second principal aggregation mechanism is through the formation of intermolecular interactions of the P,Q-rich sequences. The physico-chemical principles affecting the first growth mechanism are completely unknown but there is a natural limit to growth when all the phosphate centres have reacted to form the nanoclusters. The mechanism that limits the growth of the casein micelle by the second route has been recognised, although not fully understood, for many years (Talbot & Waugh, 1967; Waugh & Talbot, 1971), ever since  $\kappa$ -casein was first isolated and studied. Thus, the solution to the problem of indefinite association through the interaction of P,Q-rich sequences was to introduce another protein into the casein mixture, namely  $\kappa$ -casein, with the ability to bind promiscuously to the P,Q-rich sequences of the other caseins, and by so doing, terminate any further growth. We can describe the action of  $\kappa$ -casein in limiting the size of the aggregate as analogous to that of a molecular chaperone because molecular chaperones can limit precipitation of amorphous aggregates as well as controlling the growth of amyloid structures.  $\kappa$ -Casein can act as a molecular chaperone in this context because (a) it does not contain a calcium phosphate binding motif and (b) while it contains a sticky P,Q-rich sequence that allows it to bind to the other caseins, this sequence has been modified along part of its length to resemble that of a soluble, P,S,T-rich mucin (Fig. 2). In some manner which we still do not completely understand,  $\kappa$ -casein can successfully compete with the interaction of P,Q-rich sequences among the calcium-sensitive caseins even though it is usually present as a minor mole fraction of total casein.

The story does not end here. One might suppose that the P,Q-rich sequences in caseins are unnecessary. After all, a single calcium phosphate nanocluster sequestered by 50 or so acidic peptides containing a single phosphate centre can also form a thermodynamically stable

solution. Not only are the P,Q-rich sequences unnecessary from this perspective, but also they are highly troublesome because they harbour the most energetically favourable steric zipper sequences. From the mother's point of view, she would be better off producing milk from proteins that did not contain the sticky P,Q-rich sequences. The need for P,Q-rich sequences in a successful reproductive strategy based on lactation emerges when we consider and balance the interests of both mother and child. The sticky P,Q-rich sequences may be troublesome and unnecessary to the mother but useful to the neonate because they can form a casein gel or clot in his or her stomach. The clot can form either by reduction of the pH in the stomach or by cleaving off the mucin like sequence of  $\kappa$ -casein with an aspartate proteinase such as chymosin or pepsin. In both cases, the intermolecular interactions that cause the gel to form are the same type of interactions of Pro-rich sequences that have been exploited in the formation of many other types of viscous fluids and gels in biology (Williamson, 1994).

## *7. Conclusions*

It is a truism that the structure of the casein micelle is incompletely known. The same can be said of many other protein structures, particularly heterogeneous, oligomeric ones, because they are progressively refined and improved upon. The question is whether the model of the casein micelle in the future will be a refinement of one of the current range of nanocluster models or a completely different one. The possibility of the latter cannot be dismissed lightly. The structure of the casein micelle is perturbed easily by many of the methods used in its study. As the quantum physicist Manfred Eigen observed "A theory has only the alternative of being right or wrong. A model has a third possibility: it may be right, but irrelevant" (Eigen, 1973). In this context, a model based on artefactual observations, no matter how well it matches those observations, will ultimately prove irrelevant. The simplest nanocluster model, unlike its predecessors, is based primarily on observations made by solution X-ray and neutron scattering under conditions that preserved the native environment of the micelle so that the possibility of artefacts is minimised. The radius, mass and average spacing of the calcium phosphate nanoclusters, which comprise the only regular substructure of the simplest nanocluster model, have all been determined by independent experimental observations. The model establishes a clear relationship between the structure and its biological functions although, as we have discussed here, it has been realised that the range of biological functions

is broader than the original nutritional role. The model also predicts successfully the partition of salts in milk under physiological conditions. Moreover, although there is nothing closely comparable to a casein micelle in the rest of biology (so far as we know), the ability to sequester amorphous calcium phosphate and the ability to form open, highly hydrated, structures are properties that caseins share with other, closely related, unfolded proteins.

Nevertheless, even if a nanocluster model for the casein micelle or its refinements endure, a model is not enough. We also need a theory of micelle formation that establishes the physico-chemical mechanisms of growth of the particle and how that growth terminates to produce the observed structure, substructure and size distribution.

### ***Acknowledgements***

This work was funded in part by Dairy Australia. HE is supported by an Australian Research Council Future Fellowship (FT110100586). CH received an honorarium from the organisers of the IDF Symposium on Microstructure of Dairy Products, 3-7 March 2014, Melbourne, Australia.

We thank Sarah Meehan, Margaret Sunde, Tomas Koudelka, Phillip Morgan, Teresa Treweek and William Price for their considerable contributions to aspects of our casein research.

### ***8. Legends to figures***

**Fig. 1** The bauplan of casein genes. (a) Calcium-sensitive caseins, (b) calcium-insensitive,  $\kappa$ -casein, genes

**Fig. 2.** A polar plot of amino acid composition of translated sequences from different exon types in a number of eutherian, marsupial and monotreme caseins as defined in (Holt, Carver, Ecroyd, & Thorn, 2013; Kawasaki, Lafont, & Sire, 2011). Single letter amino acid codes identify each radial axis and the length of the axis is the average mol % of that residue in the sampled sequences. (a) Exons encoding signal peptides, (b) all other short exons encoding phosphate centres and flanking sequences but excluding the short exons encoding the C-terminus of  $\beta$ -caseins, (c) long exons encoding P,Q-rich sequences, (d) the N-terminal half of

the sequences encoded by long exons in  $\kappa$ -casein, (e) the C-terminal half of the sequences encoded by the long exons in  $\kappa$ -casein, (f) For comparison with Fig. 2(e), the P,S,T-rich mucin composition of human MUC-7 (<http://www.uniprot.org/uniprot/Q8TAX7>) (Bobek, Tsai, Biesbrock, & Levine, 1993).

## 9. References

- Abeln, S., & Frenkel, D. (2008). Disordered flanks prevent peptide aggregation. *Plos Computational Biol*, 4.
- Adzhubei, A. A., Sternberg, M. J. E., & Makarov, A. A. (2013). Polyproline-II helix in proteins: structure and function. *J Mol Biol*, 425, 2100-2132.
- Aoki, T., Yamada, N., Tomita, I., Kako, Y., & Imamura, T. (1987). Caseins are cross-linked through their ester phosphate groups by colloidal calcium-phosphate. *Biochim Biophys Acta*, 911, 238-243.
- Baranyi, M., Brignon, G., Anglade, P., & Ribadeau-Dumas, B. (1995). New data on the proteins of rabbit (*Oryctolagus-cuniculus*) milk. *Comp Biochem Physio B-Biochemistry & Mol Biol*, 111, 407-415.
- Barral, J. M., Broadley, S. A., Schaffar, G., & Hartl, F. U. (2004). Roles Mol chaperones in protein misfolding diseases. *Seminars in Cell & Developmental Biol*, 15, 17-29.
- Bertoncini, C. W., Jung, Y. S., Fernandez, C. O., Hoyer, W., Griesinger, C., Jovin, T. M., & Zweckstetter, M. (2005). Release of long-range tertiary interactions potentiates aggregation of natively unstructured alpha-synuclein. *Proc Nat Acad USA*, 102, 1430-1435.
- Bhattacharyya, J., & Das, K. P. (1999). Mol chaperone-like properties of an unfolded protein, alpha(s)-casein. *J Biol Chem*, 274, 15505-15509.
- Bobek, L. A., Tsai, H., Biesbrock, A. R., & Levine, M. J. (1993). Mol cloning, sequence, and specificity of expression of the gene encoding the low Mol weight human salivary mucin (MUC7). *J Biol Chem*, 268, 20563-20569.

Bouguyon, E., Beauvallet, C., Huet, J. C., & Chanut, E. (2006). Disulphide bonds in casein micelle from milk. *Biochem Biophys Res Commun*, 343, 450-458.

Boumahrou, N., Andrei, S., Miranda, G., Henry, C., Panthier, J. J., Martin, P., & Bellier, S. (2009). The major protein fraction of mouse milk revisited using proven proteomic tools. *J Physiol Pharmacol*, 60, 113-118.

Carrotta, R., Canale, C., Diaspro, A., Trapani, A., San Biagio, P. L., & Bulone, D. (2012). Inhibiting effect of alpha(S1)-casein on A beta(1-40) fibrillogenesis. *Biochim Biophys Acta*, 1820, 124-132.

Carver, J. A., Rekas, A., Thorn, D. C., & Wilson, M. R. (2003). Small heat-shock proteins and clusterin: Intra- and extracellular Mol chaperones with a common mechanism of action and function? *IUBMB Life*, 55, 661-668.

Chiti, F., & Dobson, C. M. (2006). Protein misfolding, functional amyloid, and human disease. *Ann Rev Biochem*, 75, 333-366.

Clegg, R. A., & Holt, C. (2009). An *E. coli* over-expression system for multiply-phosphorylated proteins and its use in a study of calcium phosphate sequestration by novel recombinant phosphopeptides. *Protein Expression Purif*, 67, 23-34.

Cooper, A. (2000). Heat capacity of hydrogen-bonded networks: an alternative view of protein folding thermodynamics. *Biophys Chem*, 85, 25-39.

Cooper, A. (2005). Heat capacity effects in protein folding and ligand binding: a re-evaluation of the role of water in biomolecular thermodynamics. *Biophys Chem*, 115, 89-97.

Cramer, C. J., & Truhlar, D. G. (1992). An SCF solvation model for the hydrophobic effect and absolute free energies of aqueous solvation. *Science*, 256, 213-217.

Cubellis, M. V., Caillez, F., Blundell, T. L., & Lovell, S. C. (2005). Properties of polyproline II, a secondary structure element implicated in protein-protein interactions. *Proteins-Structure Function Bioinformatics*, 58, 880-892.



Dalgleish, D. G. (2011). On the structural models of bovine casein micelles-review and possible improvements. *Soft Matter*, 7, 2265-2272.

Dehle, F. C., Ecroyd, H., Musgrave, I. F., & Carver, J. A. (2010).  $\alpha$  B-Crystallin inhibits the cell toxicity associated with amyloid fibril formation by kappa-casein and the amyloid-beta peptide. *Cell Stress & Chaperones*, 15, 1013-1026.

Djikaev, Y. S., & Ruckenstein, E. (2009). Effect of hydrogen bond networks on the nucleation mechanism of protein folding. *Physical review. E, Statistical Nonlinear, Soft Matter Phys*, 80, 061918.

Dobzhansky, T. (1973). Nothing in Biol makes sense except in the light of evolution. *Amer Biol Teacher*, 35, 125-129.

Dyson, H. J., & Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nature Rev Mol Cell Biol*, 6, 197-208.

Ecroyd, H., Koudelka, T., Thorn, D. C., 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. *J Biol Chem*, 283, 9012-9022.

Ecroyd, H., Thorn, D. C., Liu, Y., & Carver, J. A. (2010). The dissociated form of k-casein is the precursor to its amyloid fibril formation. *Biochem J*, 429, 251-260.

Ecroyd, H., Garvey, M., Thorn, D. C., Gerrard, J. A., & Carver, J. A. (2013). Amyloid fibrils from readily available sources: milk casein and lens crystallin proteins. *Methods Mol Biol*, 996, 103-117.

Eigen, M. (1973). The origin of biological information. In J. Mehra (Ed.), *The Physicist's Conception of Nature* (pp. 594-633). Dordrech, The Netherlands: D. Reidel.

Farrell, H. M., Cooke, P. H., Wickham, E. D., Piotrowski, E. G., & Hoagland, P. D. (2003). Environmental influences on bovine kappa-casein: Reduction and conversion to fibrillar (amyloid) structures. *J Protein Chem*, 22, 259-273.

Farrell, H. M., Qi, P. X., & Uversky, V. N. (2006). New views of protein structure: applications to the caseins: protein structure and functionality. In M. L. Q. Fishman, P. X. Qi & L. Wicker (Eds.), *Advan Biopolymers* (Vol. 935, pp. 52-70). Columbus, OH, USA: ACS Symposium Series.

Fisher, L. W., Torchia, D. A., Fohr, B., Young, M. F., & Fedarko, N. S. (2001). Flexible structures of SIBLING proteins, bone sialoprotein, and osteopontin. *Biochem Biophys Res Commun*, 280, 460-465.

Gagnaire, V., Pierre, A., Molle, D., & Léonil, J. (1996). Phosphopeptides interacting with colloidal calcium phosphate isolated by tryptic hydrolysis of bovine casein micelles. *J Dairy Res*, 63, 405-422.

Goldschmidt, L., Teng, P. K., Riek, R., & Eisenberg, D. (2010). Identifying the amyloids, proteins capable of forming amyloid-like fibrils. *Proc Nat Acad Sci USA*, 107, 3487-3492.

Hall, D., Hirota, N., & Dobson, C. M. (2005). A toy model for predicting the rate of amyloid formation from unfolded protein. *J Mol Biol*, 351, 195-205.

Hall, D., & Hirota, N. (2009). Multi-scale modelling of amyloid formation from unfolded proteins using a set of theory derived rate constants. *Biophys Chem*, 140, 122-128.

Halwer, M. (1954). Light-scattering study of effect of electrolytes on alpha-casein and beta-casein solutions. *Arch Biochem Biophys*, 51, 79-87.

Hartl, F. U., Bracher, A., & Hayer-Hartl, M. (2011). Molecular chaperones in protein folding and proteostasis. *Nature*, 475, 324-332.

Hassanisadi, M., Barzear, A., Yousefi, R., Dalgalarondo, M., Chobert, J.-M., Haertle, T., Saboury, A. A., & Moosavi-Movahedi, A. A. (2008). Chemometric study of the aggregation of alcohol dehydrogenase and its suppression by beta-caseins: A mechanistic perspective. *Anal Chim Acta*, 613, 40-47.

Holland, J. W., Deeth, H. C., & Alewood, P. F. (2008). Analysis of disulphide linkages in bovine kappa-casein oligomers using two-dimensional electrophoresis. *Electrophoresis*, 29, 2402-2410.

Holt, C., Davies, D. T., & Law, A. J. R. (1986). Effects of colloidal calcium phosphate content and free calcium ion concentration in the milk serum on the dissociation of bovine casein micelles. *J Dairy Res*, 53, 557-572.

Holt, C., van Kemenade, M. J. J. M., Nelson, L. S., Sawyer, L., Harries, J. E., Bailey, R. T., & Hukins, D. W. L. (1989). Composition and structure of micellar calcium-phosphate. *J Dairy Res*, 56, 411-416.

Holt, C., & Sawyer, L. (1993). Caseins as Rheomorphic Proteins - Interpretation of Primary and Secondary Structures of the Alpha-S1-Caseins, Beta-Caseins and Kappa-Caseins. *J Chem Soc-Faraday Trans*, 89, 2683-2692.

Holt, C., Wahlgren, N. M., & Drakenberg, T. (1996). Ability of a beta-casein phosphopeptide to modulate the precipitation of calcium phosphate by forming amorphous dicalcium phosphate nanoclusters. *Biochem J*, 314, 1035-1039.

Holt, C. (1998). Casein micelle substructure and calcium phosphate interactions studied by sephacryl column chromatography. *J Dairy Sci*, 81, 2994-3003.

Holt, C., Timmins, P. A., Errington, N., & Leaver, J. (1998). A core-shell model of calcium phosphate nanoclusters stabilized by beta-casein phosphopeptides, derived from sedimentation equilibrium and small-angle X-ray and neutron-scattering measurements. *Eur J Biochem*, 252, 73-78.

Holt, C., Sorensen, E. S., & Clegg, R. A. (2009). Role of calcium phosphate nanoclusters in the control of calcification. *FEBS J*, 276, 2308-2323.

Holt, C., & Carver, J. A. (2012). Darwinian transformation of a 'scarcely nutritious fluid' into milk. *J Evol Biol*, 25, 1253-1263.

Holt, C. (2013). Unfolded phosphoproteins enable soft and hard tissues to coexist in the same organism with relative ease. *Current Opinion Structural Biol*, 23, 420-425.

Holt, C., Carver, J. A., Ecroyd, H., & Thorn, D. C. (2013). Caseins and the casein micelle: their biological functions, structures and behaviour in foods. *J Dairy Sci*, 96, 6127–6146.

Holt, C., Lenton, S., Nylander, T., & Teixeira, S. C. M. (2014). Mineralisation of soft and hard tissues and the stability of biofluids. *J Structural Biol*, 185, 383–396.

Horita, D. A., Baldisseri, D. M., Zhang, W. X., Altieri, A. S., Smithgall, T. E., Gmeiner, W. H., & Byrd, R. A. (1998). Solution structure of the human Hck SH3 domain and identification of its ligand binding site. *J Mol Biol*, 278, 253-265.

Horne, D. S., Lucey, J. A., & Choi, J.-W. (2007). Casein interactions: Does the chemistry really matter? In E. Dickinson & M. E. Leser (Eds.), *Food Colloids: Self-Assembly and Material Science* (Vol. 302, pp. 155-166). London, UK: Royal Society of Chemistry Special Publications

Hoyer, W., Cherny, D., Subramaniam, V., & Jovin, T. M. (2004). Impact of the acidic C-terminal region comprising amino acids 109-140 on alpha-synuclein aggregation in vitro. *Biochemistry*, 43, 16233-16242.

Hsu, W. L., Oldfield, C. J., Xue, B., Meng, J. W., Huang, F., Romero, P., Uversky, V. N., & Dunker, A. K. (2013). Exploring the binding diversity of intrinsically disordered proteins involved in one-to-many binding. *Protein Sci*, 22, 258-273.

Jiménez, J. L., Nettleton, E. J., Bouchard, M., Robinson, C. V., Dobson, C. M., & Saibil, H. R. (2002). The protofilament structure of insulin amyloid fibrils. *Proc Nat Acad Sci USA*, 99, 9196-9201.

Kauzmann, W. (1959). Some factors in the interpretation of protein denaturation. *Advan Protein Chem*, 14, 1-63.

Kawasaki, K., & Weiss, K. M. (2003). Mineralized tissue and vertebrate evolution: The secretory calcium-binding phosphoprotein gene cluster. *Proc Nat Acad USA*, 100, 4060-4065.

Kawasaki, K. (2009). The SCPP gene repertoire in bony vertebrates and graded differences in mineralized tissues. *Dev Genes Evol*, 219, 147-157.

Kawasaki, K., Lafont, A.-G., & Sire, J.-Y. (2011). The evolution of casein genes from tooth genes before the origin of mammals. *Mol Biol Evol*, 28, 2053-2061.

Kay, B. K., Williamson, M. P., & Sudol, P. (2000). The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J*, 14, 231-241.

Khodarahmi, R., Beyrami, M., & Soori, H. (2008). Appraisal of casein's inhibitory effects on aggregation accompanying carbonic anhydrase refolding and heat-induced ovalbumin fibrillogenesis. *Arch Biochem Biophys*, 477, 67-76.

Kornberg, A. (1989). Never a dull enzyme. *Ann Rev Biochem*, 58, 1-31.

Koudelka, T., Hoffmann, P., & Carver, J. A. (2009). Dephosphorylation of aS- and b-caseins and its effect on chaperone activity: A structural and functional investigation. *J Agric Food Chem*, 57, 5956-5964.

Kresheck, G. C. (1965). Conformation of casein in aqueous solution. *Acta Chem Scand*, 19, 375-382.

Lefèvre, C. M., Sharp, J. A., & Nicholas, K. R. (2009). Characterisation of monotreme caseins reveals lineage-specific expansion of an ancestral casein locus in mammals. *Reprod FertilityDev*, 21, 1015-1027.

Lefèvre, C. M., Sharp, J. A., & Nicholas, K. R. (2010). Evolution of lactation: ancient origin and extreme adaptations of the lactation system. *Ann Rev Genomics Human Genetics*, 11, 219-238.

- Lemay, D. G., Lynn, D. J., Martin, W. F., Neville, M. C., Casey, T. M., Rincon, G., Kriventseva, E. V., Barris, W. C., Hinrichs, A. S., Molenaar, A. J., Pollard, K. S., Maqbool, N. J., Singh, K., Murney, R., Zdobnov, E. M., Tellam, R. L., Medrano, J. F., German, J. B., & Rijnkels, M. (2009). The bovine lactation genome: insights into the evolution of mammalian milk. *Genome Biol*, 10, R43.
- Léonil, J., Henry, G., Jouanneau, D., Delage, M. M., Forge, V., & Putaux, J. L. (2008). Kinetics of fibril formation of bovine kappa-casein indicate a conformational rearrangement as a critical step in the process. *J Mol Biol*, 381, 1267-1280.
- Little, E. M., & Holt, C. (2004). An equilibrium thermodynamic model of the sequestration of calcium phosphate by casein phosphopeptides. *Eur Biophys J Biophys Lett*, 33, 435-447.
- Maccallum, P. H., Poet, R., & Milner-White, E. J. (1995). Coulombic attractions between partially charged main-chain atoms stabilize the right-handed twist found in most beta-strands. *J Mol Biol*, 248, 374-384.
- Macias, M. J., Wiesner, S., & Sudol, M. (2002). WW and SH3 domains, two different scaffolds to recognize proline-rich ligands. *FEBS Lett*, 513, 30-37.
- Martin-Garcia, J. M., Ruiz-Sanz, J., & Luque, I. (2012). Interfacial water molecules in SH3 interactions: a revised paradigm for polyproline recognition. *BiochemJ*, 442, 443-451.
- Martin, P., Ferranti, P., Leroux, C., & Addeo, F. (2003). Non-bovine caseins: qualitative variability and Mol diversity. In P. F. Fox & P. L. H. McSweeney (Eds.), *Advanced Dairy Chemistry* (3rd ed., Vol. 1A Proteins Basic Aspects, pp. 277-317). New York, NY, USA: Kluwer Academic/Plenum.
- Martin, P., Cebo, C., & Miranda, G. (2013). Interspecies comparison of milk proteins: quantitative variability and Mol diversity. In P. L. H. McSweeney & P. F. Fox (Eds.), *Advanced Dairy Chemistry* (4th ed., Vol. 1A. Proteins: Basic Aspects, pp. 386-429). New York, NY, USA: Springer.

- Matsudomi, N., Kanda, Y., Yoshika, Y., & Moriwaki, H. (2004). Ability of alpha-S-casein to suppress the heat aggregation of ovotransferrin. *J Agric Food Chem*, 52, 4882-4886.
- Mazzali, M., Kipari, T., Ophascharoensuk, V., Wesson, J. A., Johnson, R., & Hughes, J. (2002). Osteopontin - a molecule for all seasons. *QJM Intl J Medicine*, 95, 3-13.
- McGann, T. C. A., & Pyne, G. T. (1960). The colloidal phosphate of milk 3. Nature of its association with casein. *J Dairy Res*, 27, 403-417.
- McMeekin, T. L. (1952). Milk Proteins. *J Food Protection*, 15, 57-63.
- McMeekin, T. L. (1970). Milk proteins in retrospect. In H. A. McKenzie (Ed.), *Milk proteins chemistry and Mol Biol* (Vol. 1, pp. 3-15). New York, NY, USA and London, UK: Academic Press.
- Mezei, M., Fleming, P. J., Srinivasan, R., & Rose, G. D. (2004). Polypyrrolone II helix is the preferred conformation for unfolded polyalanine in water. *Proteins-Structure Function and Bioinformatics*, 55, 502-507.
- Mikheeva, L. M., Grinberg, N. V., Grinberg, V. Y., Khokhlov, A. R., & de Kruif, C. G. (2003). Thermodynamics of micellization of bovine beta-casein studied by high-sensitivity differential scanning calorimetry. *Langmuir*, 19, 2913-2921.
- Miranda, G., Mahe, M. F., Leroux, C., & Martin, P. (2004). Proteomic tools to characterize the protein fraction of Equidae milk. *Proteomics*, 4, 2496-2509.
- Moffatt, P., Smith, C. E., Sooknanan, R., St-Arnaud, R., & Nanci, A. (2006). Identification of secreted and membrane proteins in the rat incisor enamel organ using a signal-trap screening approach. *Eur J Oral Sci*, 114, 139-146.
- Morgan, P. E., Treweek, T. M., Lindner, R. A., Price, W. E., & Carver, J. A. (2005). Casein proteins as Mol chaperones. *J Agric Food Chem*, 53, 2670-2683.

Murphy, C. L., Kestler, D. P., Foster, J. S., Wang, S., Macy, S. D., Kennel, S. J., Carlson, E. R., Hudson, J., Weiss, D. T., & Solomon, A. (2008). Odontogenic ameloblast-associated protein nature of the amyloid found in calcifying epithelial odontogenic tumors and unerupted tooth follicles. *Amyloid-JProtein Folding Disorders*, 15, 89-95.

Nelson, R., Sawaya, M. R., Balbirnie, M., Madsen, A. O., Riek, C., Grothe, R., & Eisenberg, D. (2005). Structure of the cross- $\beta$  spine of amyloid-like fibrils. *Nature*, 435, 773-778.

Niewold, T. A., Murphy, C. L., Hulskamp-Koch, C. A. M., Tooten, P. C. J., & Gruys, E. (1999). Casein related amyloid, characterization of a new and unique amyloid protein isolated from bovine corpora amylacea. *Amyloid-International J Exper Clin Invest*, 6, 244-249.

Ofstedal, O. T. (2012). The evolution of milk secretion and its ancient origins. *Animal*, 6, 355-368.

Ono, T., Ohtawa, T., & Takagi, Y. (1994). Complexes of casein phosphopeptides and calcium phosphate prepared from casein micelles by tryptic digestion. *Biosci Biotechnol Biochem*, 58, 1376-1380.

Payens, T. A. J., & Vreeman, H. J. (1982). Casein micelles and micelles of  $\kappa$ - and  $\beta$ -casein. In K. L. Mittal (Ed.), *Solution Behaviour of Surfactants* (Vol. 1, pp. 543-571). New York, NY, USA: Plenum.

Perutz, M. F., Finch, J. T., Berriman, J., & Lesk, A. (2002). Amyloid fibers are water-filled nanotubes. *Proc Nat Acad USA*, 99, 5591-5595.

Rasmussen, L. K., Due, H. A., & Petersen, T. E. (1995). Human  $\alpha$ (S1)-casein-purification and characterization. *Comp Biochem Physiol B*, 111, 75-81.

Rasmussen, L. K., Johnsen, L. B., Tsiora, A., Sorensen, E. S., Thomsen, J. K., Nielsen, N. C., Jakobsen, H. J., & Petersen, T. E. (1999). Disulphide-linked caseins and casein micelles. *Int Dairy J*, 9, 215-218.



- Raynes, J. K., & Gerrard, J. A. (2013). Amyloid Fibrils as Bionanomaterials. In B. H. A. Rehm (Ed.), *Bionanotechnology: Biological Self-Assembly and its Applications* (pp. 85-106). Wymondham, UK: Caister Academic Press.
- Rijnkels, M., Kooiman, P. M., de Boer, H. A., & Pieper, F. R. (1997). Organization of the bovine casein gene locus. *Mamm Genome*, 8, 148-152.
- Rose, G. D. (2002). Unfolded Proteins. In F. M. Richards, D. S. Eisenberg & J. Kuriyan (Eds.), *Advan Protein Chem* (Vol. 62, pp. 398). Amsterdam, The Netherlands: Academic Press.
- Sawaya, M. R., Sambashivan, S., Nelson, R., Ivanova, M. I., Sievers, S. A., Apostol, M. I., Thompson, M. J., Balbirnie, M., Wiltzius, J. J. W., McFarlane, H. T., Madsen, A. O., Riek, C., & Eisenberg, D. (2007). Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature*, 447, 453-457.
- Scatena, M., Liaw, L., & Giachelli, C. M. (2007). Osteopontin - A multifunctional molecule regulating chronic inflammation and vascular disease. *Arteriosclerosis Thrombosis and Vascular Biol*, 27, 2302-2309.
- Shewry, P. R., & Halford, N. G. (2002). Cereal seed storage proteins: structures, properties and role in grain utilization. *J Exper Botany*, 53, 947-958.
- Shi, Z. S., Chen, K., Liu, Z. G., & Kallenbach, N. R. (2006). Conformation of the backbone in unfolded proteins. *Chem Rev*, 106, 1877-1897.
- Siligardi, G., Ruzza, P., Hussain, R., Cesaro, L., Brunati, A. M., Pinna, L. A., & Donella-Deana, A. (2012). The SH3 domain of HS1 protein recognizes lysine-rich polyproline motifs. *Amino Acids*, 42, 1361-1370.
- Stapley, B. J., & Creamer, T. P. (1999). A survey of left-handed polyproline II helices. *Protein Sci*, 8, 587-595.

Syme, C. D., Blanch, E. W., Holt, C., Jakes, R., Goedert, M., Hecht, L., & Barron, L. D. (2002). A Raman optical activity study of rheomorphism in caseins, synucleins and tau - new insight into the structure and behaviour of natively unfolded proteins. *Eur J Biochem*, 269, 148-156.

Syme, N. R., Dennis, C., Phillips, S. E. V., & Homans, S. W. (2007). Origin of heat capacity changes in a "Nonclassical" hydrophobic interaction. *Chembiochem*, 8, 1509-1511.

Talbot, B., & Waugh, D. F. (1967). Coat-core interactions in casein micelles. *J Dairy Sci*, 50, 949-&.

Thorn, D. C., Meehan, S., Sunde, M., Rekas, A., Gras, S. L., MacPhee, C. E., Dobson, C. M., Wilson, M. R., & Carver, J. A. (2005). Amyloid fibril formation by bovine milk kappa-casein and its inhibition by the Mol chaperones alpha(s-) and beta-casein. *Biochemistry*, 44, 17027-17036.

Thorn, D. C., Ecroyd, H., Sunde, M., Poon, S., & Carver, J. A. (2008). Amyloid fibril formation by bovine milk alpha(s2)-casein occurs under physiological conditions yet is prevented by its natural counterpart, alpha(s1)-casein. *Biochemistry*, 47, 3926-3936.

Tompa, P. (2012). Intrinsically disordered proteins: a 10-year recap. *Trends Biochem Sci*, 37, 509-516.

Treweek, T. M., Thorn, D. C., Price, W. E., & Carver, J. A. (2011). The chaperone action of bovine milk alpha(S1)- and alpha(S2)-caseins and their associated form alpha(S)-casein. *Arch Biochem Biophys*, 510, 42-52.

Uversky, V. N., Gillespie, J. R., & Fink, A. L. (2000). Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins-Structure Function and Genetics*, 41, 415-427.

Uversky, V. N., & Dunker, A. K. (2010). Understanding protein non-folding. *Biochim Biophys Acta*, 1804, 1231-1264.

- Wald, T., Osickova, A., Sulc, M., Benada, O., Semeradtova, A., Rezabkova, L., Veverka, V., Bednarova, L., Maly, J., Macek, P., Sebo, P., Slaby, I., Vondrasek, J., & Osicka, R. (2013). Intrinsically disordered enamel matrix protein ameloblastin forms ribbon-like supraMol structures via an N-terminal segment encoded by exon 5. *J Biol Chem*, 288, 22333-22345.
- Waugh, D. F., & Talbot, B. (1971). Equilibrium casein micelle systems. *Biochemistry*, 10, 4153-4162.
- Williamson, M. P. (1994). The structure and function of proline-rich regions in proteins. *Biochem J*, 297, 249-260.
- Wiltzius, J. J. W., Sievers, S. A., Sawaya, M. R., Cascio, D., Popov, D., Riek, C., & Eisenberg, D. (2008). Atomic structure of the cross- $\beta$  spine of islet amyloid polypeptide (amylin). *Protein Sci*, 17, 1467-1474.
- Yong, Y. H., & Foegeding, E. A. (2010). Caseins:utilizing Mol chaperone properties to control protein aggregation in foods,. *J Agric Food Chem*, 58, 685-693.
- Zhang, X. F., Fu, X. M., Zhang, H., Liu, C., Jiao, W. W., & Chang, Z. Y. (2005). Chaperone-like activity of beta-casein. *Int J Biochem Cell Biol*, 37, 1232-1240.

## 10. Figures

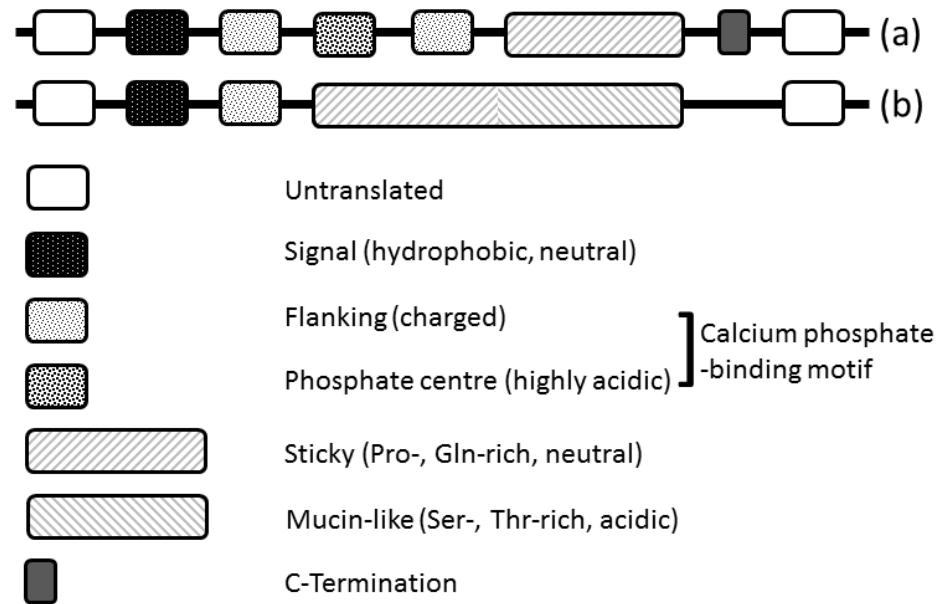


Fig. 1 The bauplan of casein genes. (a) Calcium-sensitive caseins, (b) calcium-insensitive,  $\kappa$ -casein, genes

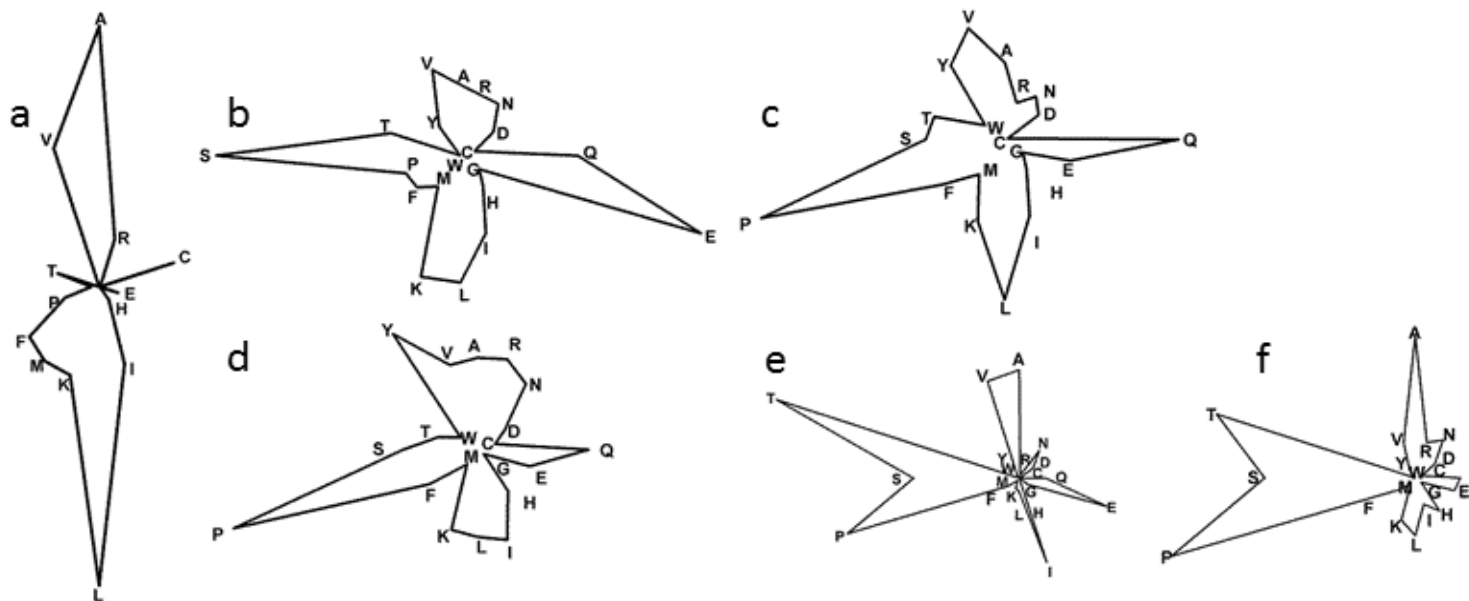


Fig. 2. A polar plot of amino acid composition of translated sequences from different exon types in a number of eutherian, marsupial and monotreme caseins as defined in (Holt, Carver, Ecroyd, & Thorn, 2013; Kawasaki, Lafont, & Sire, 2011). Single letter amino acid codes identify each radial axis and the length of the axis is the average mol % of that residue in the sampled sequences. (a) Exons encoding signal peptides, (b) all other short exons encoding phosphate centres and flanking sequences but excluding the short exons encoding the C-terminus of  $\beta$ -caseins, (c) long exons encoding P,Q-rich sequences, (d) the N-terminal half of the sequences encoded by long exons in  $\kappa$ -casein, (e) the C-terminal half of the sequences encoded by the long exons in  $\kappa$ -casein, (f) For comparison with Fig. 2(e), the P,S,T-rich mucin composition of human MUC-7 (<http://www.uniprot.org/uniprot/Q8TAX7> (Bobek, Tsai, Biesbrock, & Levine, 1993)).