Are casein micelles extracellular condensates formed by liquid-liquid phase separation?

Attila Horvath[1], Monika Fuxreiter[2], Michele Vendruscolo[3], Carl Holt[4], John A. Carver[5,*]

[1]*John Curtin School of Medical Research, The Australian National University, Acton, ACT 2601, Australia*

[2]*Department of Biomedical Sciences, University of Padova, Via Ugo Bassi, 58/B 35131, Padova, Italy*

[3]*Centre for Misfolding Diseases, Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom*

[4]*Institute of Molecular, Cell and Systems Biology, University of Glasgow, Glasgow G12 8QQ, United Kingdom*

[5]*Research School of Chemistry, The Australian National University, Acton, ACT 2601, Australia*

*To whom correspondence should be addressed at:

Phone: +61 2 6125 9748

e-mail: john.carver@anu.edu.au

Abstract

Casein micelles are extracellular polydisperse assemblies of unstructured casein proteins, which are the major component of milk. Within casein micelles, casein molecules are stabilised by binding to calcium phosphate nanoclusters and, by acting as molecular chaperones, through multivalent interactions. In light of such interactions, we discuss whether casein micelles can be considered as extracellular condensates formed by liquid-liquid phase separation. We analyse the sequence, structure and interactions of caseins in comparison to proteins forming intracellular condensates. Furthermore, we review the similarities between caseins and small heat-shock proteins whose chaperone activity is linked to phase separation of proteins. By bringing these observations together, we describe a regulatory mechanism for protein condensates, as exemplified by casein micelles.

Keywords

Casein, casein micelle, liquid-liquid phase separation, biomolecular condensate, multivalent interactions, nanocluster, molecular chaperone, protein aggregation

Abbreviations

CaP, calcium phosphate; DPR, droplet-promoting region; LARKS, low-complexity aromatic-rich kinked segments; OPN, osteopontin; SCPP, secreted calcium- or calcium phosphate-binding phosphoprotein; SLiM, short linear motif

Introduction: Caseins and casein micelles

In mammalian milk, caseins are the predominant proteins, for example making up about 80% of the protein content in cow milk. Caseins are best known for their hyper-expression in mammary tissue, where they have a role in the prevention of pathological calcification and amyloidosis, but they have a wide tissue distribution and a range of biological functions beside nutrition. Caseins are secreted calcium- or calcium phosphate-binding phosphoproteins (SCPPs). There are four main cow caseins: α_{S1} (40%), α_{S2} (12%), β (36%) and κ (11%), with their approximate percentage composition given in brackets. Caseins are secreted proteins with N-terminal signal sequences of 15 residues for α_{S1}, α_{S2} and β, and 21 residues for κ-casein. Mature caseins have a molecular mass of 19-25 kDa, with little sequence similarity but similar amino acid compositions. All four main caseins have P,Q-rich polar tract sequences (i.e., sequences of mainly polar but uncharged residues [1, 2]). Three of the caseins (α_{S1}, α_{S2} and β) have phosphorylated, calcium phosphate (CaP)-binding short linear sequence motifs (CaP-SLiMs) such as 15-pSLpSpSpSEE-21 of β-casein, where pS is phosphoserine (residue numbers are given for the mature proteins after proteolytical cleavage of the signal peptide). The fourth, κ-casein, has a polar tract in the N-terminal half with a soluble-mucin-like C-terminal region [3]. The amino acid composition of caseins, particularly P,Q-rich regions, predisposes them to adopt the polyproline-II secondary structure as observed experimentally for β- and κ-casein [4]. Consistent with this type of structure, and on the basis of their amino acid sequence and composition, caseins are predicted to be intrinsically disordered [5-7], as summarised in Figure 4c of [6]. Experimentally, there is evidence for an open and unfolded conformation for caseins [3, 6]. Overall, caseins are polar proteins; for the same net charge, caseins exhibit an average hydrophobicity lower than that of the great majority of globular proteins [8].

In common with many intrinsically disordered proteins, caseins have a propensity to associate to form homotypic and heterotypic assemblies [6, 9]. Their hydration, conformational flexibility and dynamics are largely retained upon self-assembly. In milk, casein is mostly bound to nanoclusters of amorphous CaP, and associate through protein-protein interactions to form a stable, polydisperse, heterogeneous and amorphous particles known as casein micelles. Typically, a casein micelle has a spherical shape with a radius of approximately 70 nm containing around 10,000 casein molecules. Because of their size, casein micelles scatter visible light in all directions leading to the white appearance of milk, even after removal of fat globules.

The application to caseins of the concepts developed for intrinsically disordered proteins has led to a detailed and multivalent binding model of casein micelles that accounts accurately for their size, sub-structure, polydispersity and composition [3, 6, 10-14]. According to these studies, about 70% of the casein molecules within casein micelles are bound directly to the CaP nanoclusters through their CaP-SLiMs, whereas the rest are considered to be free of direct linkages. As a result, casein-casein interactions in the casein micelles are formed by multivalent, heterogeneous interactions among bound and free caseins (**Figure 1** [11]).

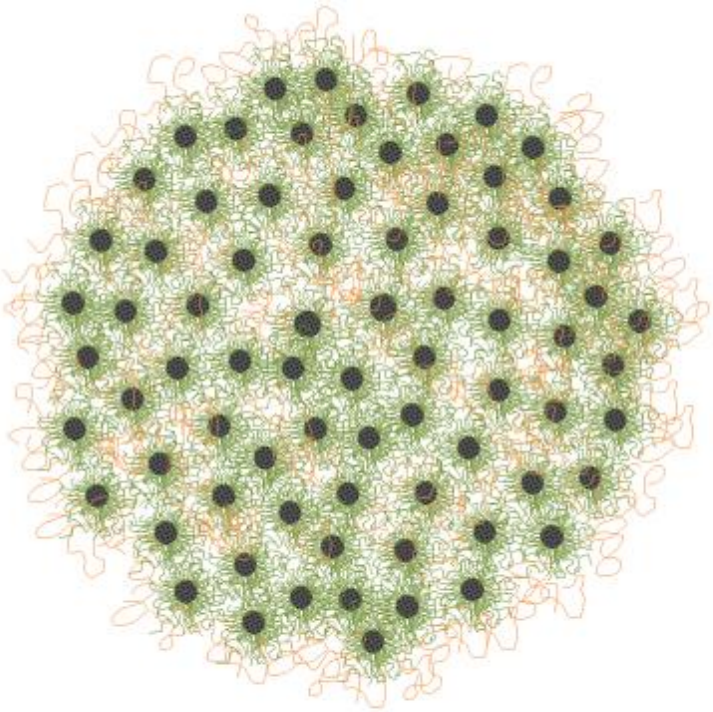

Figure 1. A schematic illustration of the core-coat structure of a typical casein micelle [11]. *The black dots represent CaP nanoclusters. The green lines represent casein molecules bound to CaP nanoclusters via highly phosphorylated sequences (CaP-SLiMs). Bound casein proteins are found in the core of the micelle, while free caseins (brown lines) are present in both the core and coat of the micelle. As the hydrated and disordered caseins interact with each other via many transient, multivalent interactions, casein micelles are dynamic entities.*

Many intrinsically disordered proteins, when at high concentrations, can form highly ordered amyloid fibrils in which the polypeptide chains are arranged in a cross β-sheet array. Amyloid fibril formation by peptides and proteins is associated with a wide range of human disorders including Alzheimer's and Parkinson's diseases [15, 16]. *In vitro*, solutions of purified cow κ-casein or α_{S2}-casein readily form amyloid fibrils under physiological conditions [17-19], whereas the other two purified caseins require harsher conditions to do so [20]. Given the intrinsic tendency of caseins to form amyloid fibrils, it is remarkable that the action of caseins as molecular chaperones prevents fibril formation from occurring within the casein micelles [21, 22].

In addition to their native and amyloid states, it has been recently demonstrated that most intracellular proteins can naturally populate a dense liquid condensate, termed the droplet state [23], which is formed by a liquid-liquid phase separation process. Liquid-liquid phase separation can give rise to non-membrane bound cellular compartments, termed membraneless organelles [24, 25]. Here, we discuss how casein association, in particular within extracellular casein micelles, may occur by a similar liquid-liquid phase separation

process to form a dispersed droplet state with comparable properties to intracellular membraneless organelles.

Liquid-liquid phase separation and membraneless organelles

Phase separation of proteins and nucleic acids leads to the formation of membraneless organelles, also termed liquid droplets or biomolecular condensates, in the nucleus and cytoplasm of cells [23, 25-27]. Membraneless organelles include well-known cellular organelles, such as nucleoli, paraspeckles, Cajal bodies, PML bodies, P-bodies and stress granules. Many membraneless organelles have been discovered recently, suggesting that others are yet to be identified [28]. The formation of membraneless organelles enables the cell to maintain a readily accessible source of protein components when needed for the maintenance of cellular homeostasis, including under conditions of cellular stress. As an example, stress granules assemble RNA and molecular chaperones under cellular stress, probably as a means of protecting RNA molecules [29-31].

Multivalent interactions between intrinsically disordered proteins and the formation of membraneless organelles

The dense, liquid-like state within membraneless organelles requires many binding configurations of the component proteins, which is achieved by multiple interaction motifs, a phenomenon known as multivalency [32]. Multivalency is exemplified by the interactions between the amino acid residues in low-complexity protein regions, i.e., ones with amino acid composition differing from that of globular proteins [23, 25]. Low-complexity regions are characterised by a high abundance of uncharged, polar amino acids such as Asn, Gln, Tyr and Ser, along with Gly. Low-complexity regions of a protein mediate phase separation to produce the droplet state, which can in turn progress to amyloid fibril formation [23, 33].

Intrinsically disordered proteins can interact with each other via multivalent interactions (**Figure 2**), for example via SLiMs [5, 6, 34-37]. These interactions may be established using multiple contact sites, which can exchange with each other in the bound complex [38]. Fuzziness is a feature of these complexes, as their conformations and interaction properties can adapt to the cellular context and thus can change with cellular conditions [9, 39]. The association of the intrinsically disordered caseins, for example, to form casein micelles, can take place via multiple contact sites of different sequence motifs and thus can be described by multivalency [11, 14]. Along these lines, the interactions between the caseins within the casein micelle are multivalent and involve SLiMs. The four caseins contain numerous SLiMs of at least four types: CaP-binding, basic, hydrophobic and order-promoting (HO) and amyloid-forming steric zippers (**Figure 2**), as outlined in our previous publications [6, 11, 12].

The dynamic and disordered nature of caseins, along with their multivalent interactions, enable caseins within micelles (particularly the 'free caseins' that are not involved in binding to CaP nanoclusters via CaP-binding SLiMs) to exchange with their environment, the milk serum. Various studies have shown that free caseins readily exchange with the serum and undergo a population shift between the casein micelles and serum in response to alteration of temperature and pressure [40, 41]. Similarly, the components (proteins, RNA, etc.) within intracellular condensates and membraneless organelles undergo constant exchange with the surrounding milieu.

Figure 2

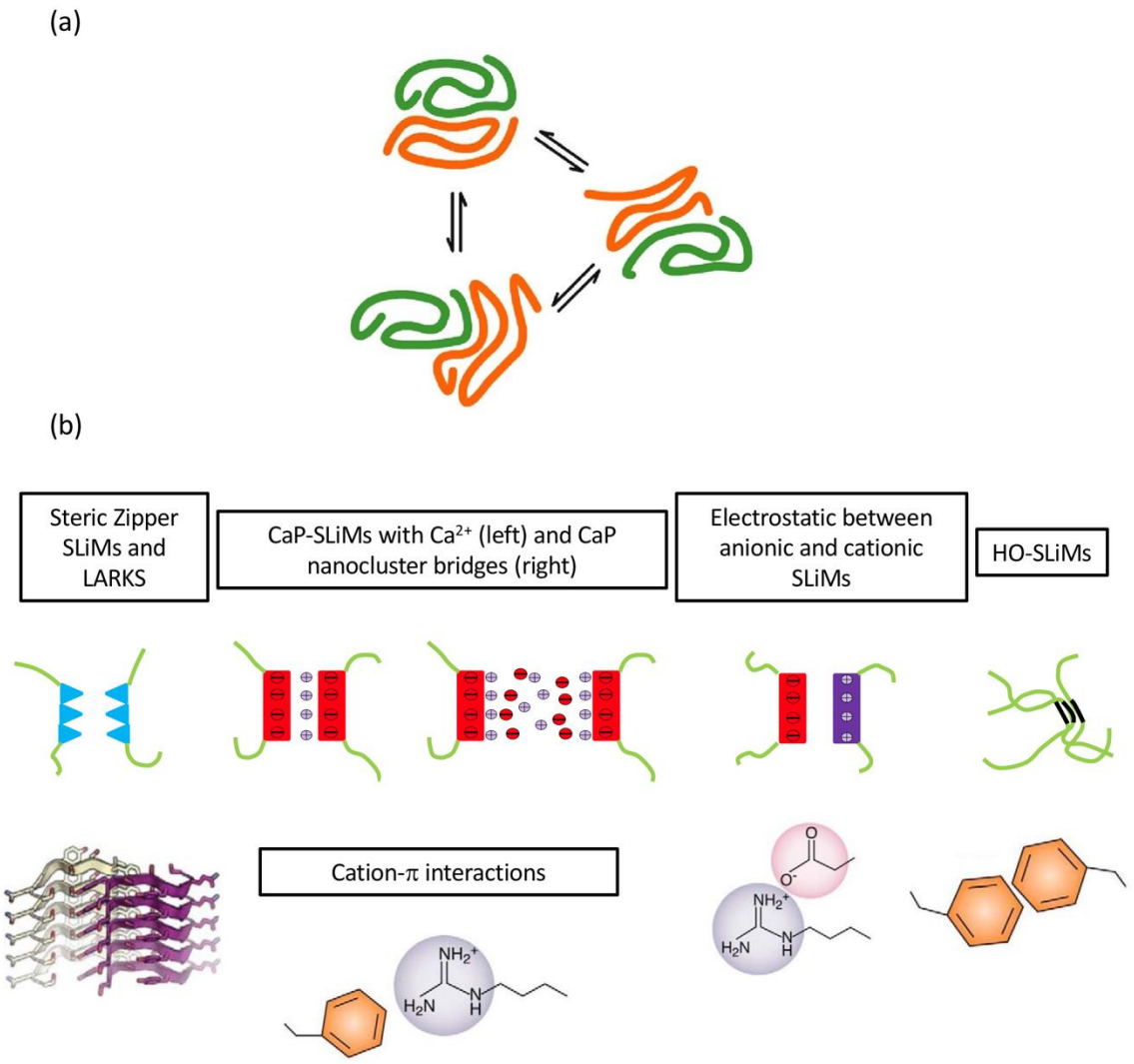

Figure 2. Examples of multivalent interactions between proteins*. (a) Schematic of dynamic multivalent interactions between a pair of intrinsically disordered proteins, such as two casein proteins within a casein micelle (with permission from [42]). Numerous weak, transient, short-range interactions facilitate oligomerisation during phase separation of proteins [25, 27, 32]. The interactions are cooperative and counteract the loss of entropy due to association of the proteins. (b) Schematic representation of the weak interactions between intrinsically disordered proteins. Left to right: Caseins contain many steric zipper SLiMs, i.e., amyloid fibril-forming hexapeptide segments, consistent with the ability of individual caseins to form stable amyloid fibrils [6, 17-20]. Low-complexity aromatic-rich kinked segments (LARKS) adopt transient cross β-sheet structures (the characteristic structural feature of amyloid fibrils) that can interact with each other [43, 44]. No LARKS have been identified in cow caseins. Neutralisation of charges via the arrangement of charged residues into stretches of amino*

acids, for example, negatively charged phosphoserine and Glu residues in CaP-SLiMs that bind Ca^{2+} and CaP in caseins [6], and electrostatic interactions between oppositely charged anionic and cationic (basic) SLiMs. HO-SLiMs via π–π interactions (stacking) between amino acids with delocalised π electrons, i.e., Tyr, Trp, Phe, Arg, Gln, Asn, Asp and Glu. There are a limited number of these interactions in cow caseins as there are only a few conserved Tyr and Phe residues within HO-SLiMs [6, 8]. Cation-π interactions occur between positively charged lysine and arginine amino acids and electron-rich aromatic amino acids. The figure is a combination of figures from [11] and [27], with permission.

Propensity of caseins to undergo liquid-liquid phase separation

The propensity of a protein to participate in multivalent interactions, and therefore be stabilised by conformational entropy, can be quantified by the FuzDrop method (https://fuzdrop.bio.unipd.it/predictor) [45]. This algorithm predicts droplet-promoting regions based on the probability of disordered interactions [46, 47] combined with the disorder score computed by the ESpritz NMR algorithm [48]. Using this approach, the probability to undergo liquid-liquid phase separation can be estimated in both globular and disordered proteins. Hardenberg *et al.* [45] used FuzDrop to analyse the propensity of sequences in the human proteome to drive phase separation as well as to form the droplet state, as present within membraneless organelles. Droplet-forming or droplet-driving sequences are enriched in the disorder-promoting amino acids, Pro, Gly and Ser, and depleted in the order-promoting, hydrophobic amino acids, Phe, Ile, Val, Cys and Trp, compared to sequences that do not undergo liquid-liquid phase separation. Droplet-forming proteins also contain elevated levels of Asn and Gln. In this context, the amino acid composition of droplet-forming sequences is in between that of globular and intrinsically disordered proteins. The average droplet-driving potential or probability (p_{LLPS}) of a protein can be determined from FuzDrop along with the droplet-driving potential of individual residues and sub-sequences [45]. It has been suggested that the spontaneous liquid-liquid phase separation of a protein can occur when its p_{LLPS} value exceeds the threshold value of 0.61 [45].

From a survey of casein sequences in 20 species, their amino acid composition is invariably of low complexity [49]. Also, except for α_{S2}-caseins, the abundance of hydrophobic amino acids is between that of typical globular proteins and typical intrinsically disordered proteins [8], suggesting that α_{S1}-, β- and κ-caseins might undergo liquid-liquid phase separation. Indeed, the average p_{LLPS} value of these unphosphorylated mature cow sequences (i.e., without their signal peptides) is above the p_{LLPS} threshold, particularly so for α_{S1}-casein (p_{LLPS} = 0.90) (**Table 1**). When the sequences are examined, two droplet-promoting regions (DPRs) of more than 25 residues are found in unphosphorylated α_{S1}-casein, and one DPR is present in unphosphorylated κ-casein. Thus, the p_{LLPS} values and identification of DPRs are consistent with the potential ability of these three unphosphorylated caseins to undergo phase separation on their own.

Effects on liquid-liquid phase separation of post-translational phosphorylation of caseins and binding to calcium phosphate nanoclusters

Post-translational modifications, such as phosphorylation, affect liquid-liquid phase separation by increasing or decreasing the hydrophobicity and valency of a protein, and hence the possibility of multivalent interactions [23]. The effect of phosphorylation on liquid-liquid phase separation of caseins was assessed by comparing the p_{LLPS} values of the unphosphorylated sequences with sequences in which the main phosphorylation sites were changed to Glu as a phosphomimic. The cow α_{S1}- and α_{S2}-caseins occur as a range of phosphoforms but, for the present purpose, the calculations were limited to the most common phosphoforms, having 8, 11, 5 and 1 phosphoserines in α_{S1}-, α_{S2}-, β- and κ-casein, respectively. In considering the effects of phosphorylation on the p_{LLPS} value of caseins, a distinction needs to be drawn between the free and bound caseins in casein micelles. The casein sequences that are phosphorylated and bound directly to the CaP nanoclusters will have their ability to participate in multivalent protein-protein interactions altered, which may affect LLPS of different caseins. To mimic the effect of binding to the CaP nanoclusters and the resultant charge neutralisation, the main sites of phosphorylation at Ser residues were changed to Ala [50] and p_{LLPS} values were calculated.

The p_{LLPS} values of the unphosphorylated and the phosphomimic forms of the cow caseins are compared in **Table 1**. The data indicate that the addition of the negative charges associated with Ser phosphorylation has little effect on the propensity for α_{S1}-, β-, and κ-caseins to undergo phase separation. The effect on α_{S2}-casein is relatively more pronounced, but this protein is the only unphosphorylated casein to have a p_{LLPS} value below the threshold. The Ala substitutions at the main sites of phosphorylation lead to a reduction in p_{LLPS} values for all four caseins which is largest for α_{S2}-casein (**Table 1**). The low p_{LLPS} value of α_{S2}-casein (**Table 1**) is in contrast to the p_{LLPS} values for the other caseins and implies that α_{S2}-casein does not phase separate. However, it could do so via the client mechanism [45] by interacting with the other caseins within the casein micelle. The client mechanism is enabled by DPRs, which can form disordered interactions.

The droplet-promoting propensity profiles (p_{DP} values versus amino acid sequence number) of the casein phosphomimics are presented in **Figure 3**. P,Q-rich sequences in intrinsically disordered proteins often participate in protein-protein interactions [10]. We previously noted [10] that the P,Q-rich sequences in caseins are polar tracts, which impart conformational flexibility to caseins. The interactions between polar tract sequences have low sequence specificity. Polar tract sequences in caseins are defined by the start and end points of the longer exons encoding them: L149-W199 in α_{S1}-casein, N83-L123 and K165-V204 in α_{S2}-casein, D43-I208 in β-casein and C11-V169 in κ-casein. The polar tract in κ-casein comprises the N-terminal P,Q-rich region (C11-S104) and F105-V169 (the so-called macropeptide) which is mucin-like in containing many hydroxylated serine and threonine residues. Of the DPRs, the long sequence Q172-W199 in α_{S1}-casein, the last four of the five sequences in β-casein (wholly or partly) and T94-M95 in κ-casein are located in the P,Q-rich sequences of each protein. In κ-casein, a DPR (P110-E140) is present within its macropeptide region.

We have previously reported that the amyloid fibril core regions of κ-casein (Y25-P86), α_{S2}-casein (A81-K113) and β-casein (Y114-V209) are contained within their respective P,Q-rich sequences [20, 51, 52]. The fibril core region of α_{S1}-casein has not been determined. For α_{S2}-

, β- and κ-casein, there is little correlation between the fibril core regions and DPRs (**Figure 3**). Similar behaviour is observed for α-synuclein, the protein whose amyloid fibril formation is intimately associated with Parkinson's disease. The disorder-promoting propensity profile of α-synuclein predicts that the disordered polar C-terminal region drives droplet formation entirely (Figure 4B of [45]), yet this region is not part of the fibril core of the protein [53]. In contrast to α-synuclein, the disorder-promoting propensity profile of α_{S1}-casein reveals that its DPRs are scattered throughout the amino acid sequence of α_{S1}-casein (**Figure 3a**). The scattering of DPRs through the sequences also occurs for β- and κ-casein (**Figures 3c, d**). In principle, this occurrence would facilitate multivalent interactions between the caseins across their sequences and hence contribute to contact site redundancy during droplet and casein micelle formation.

In a similar manner, there is good correlation between the fibril core regions in α_{S2}-, β- and κ-casein and the aggregation-promoting hotspots within their mature amino acid sequences, as identified by the CamSol method [54]. In all these three caseins, the regions with the lowest intrinsic solubility are present within the experimentally determined fibril cores (**Figure 4**), and, in the main, these regions do not coincide with the DPRs.

Figure 3

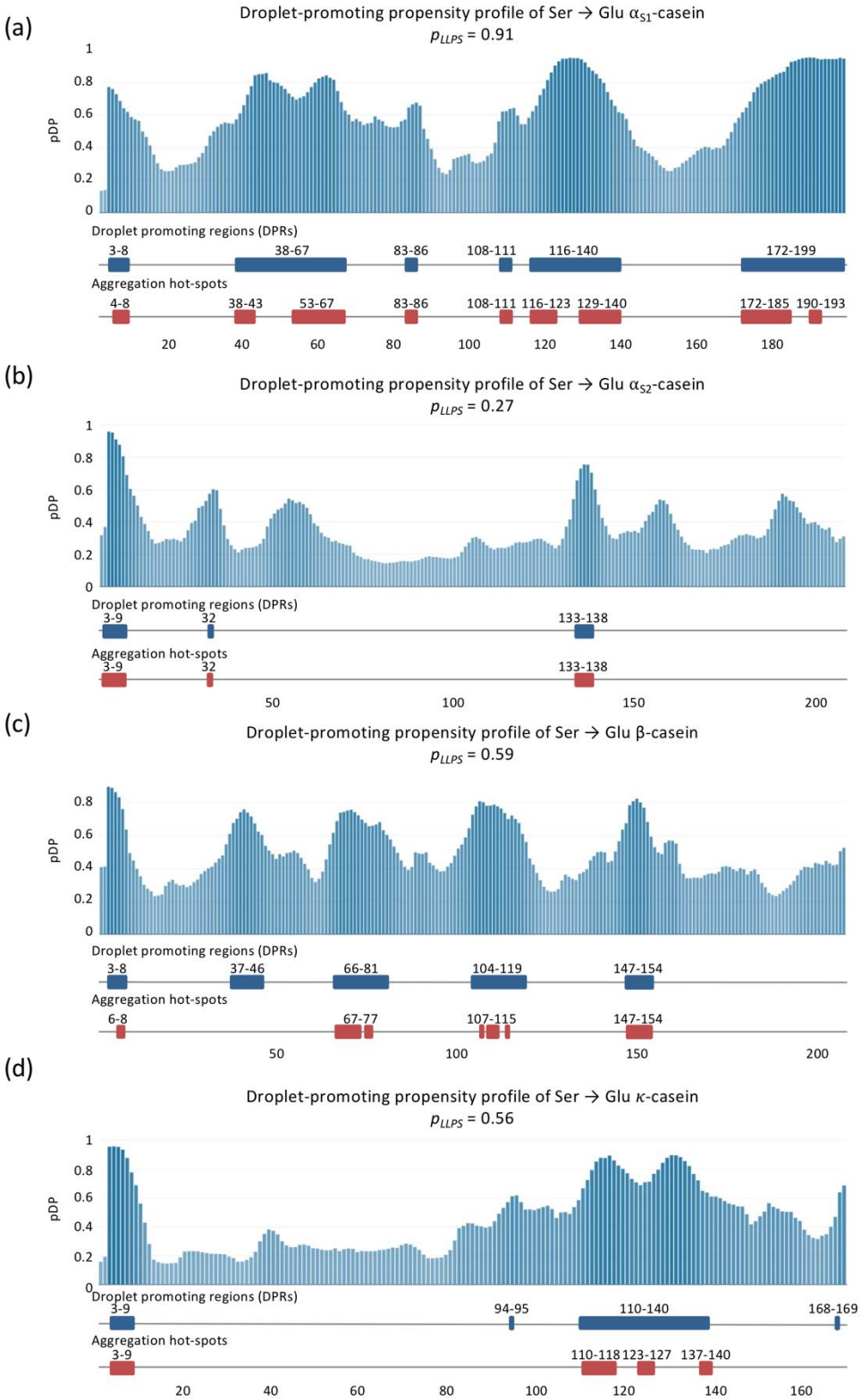

Figure 3. Droplet-promoting propensity profiles (pDP values versus amino acid sequence) of cow casein phosphomimics (Ser to Glu). *(a) α_{S1}-casein with eight phosphoserine-mimic*

residues at positions 46, 48, 64, 66, 67, 68, 75 and 115. (b) α_{S2}-casein with 11 phosphoserine-mimic residues at positions 8, 9, 10, 16, 56, 57, 58, 61, 129, 131 and 143. (c) β-casein with five phosphoserine-mimic residues at positions 15, 17, 18, 19 and 35. (d) κ-casein with one phosphoserine-mimic residue at position 149. The droplet-promoting regions encompass residues above the pLLPS threshold of 0.61. Aggregation hot-spots were identified by FuzDrop which also predicts the propensity of aggregation based on droplet-promoting propensity and binding mode entropy [33].

Figure 4

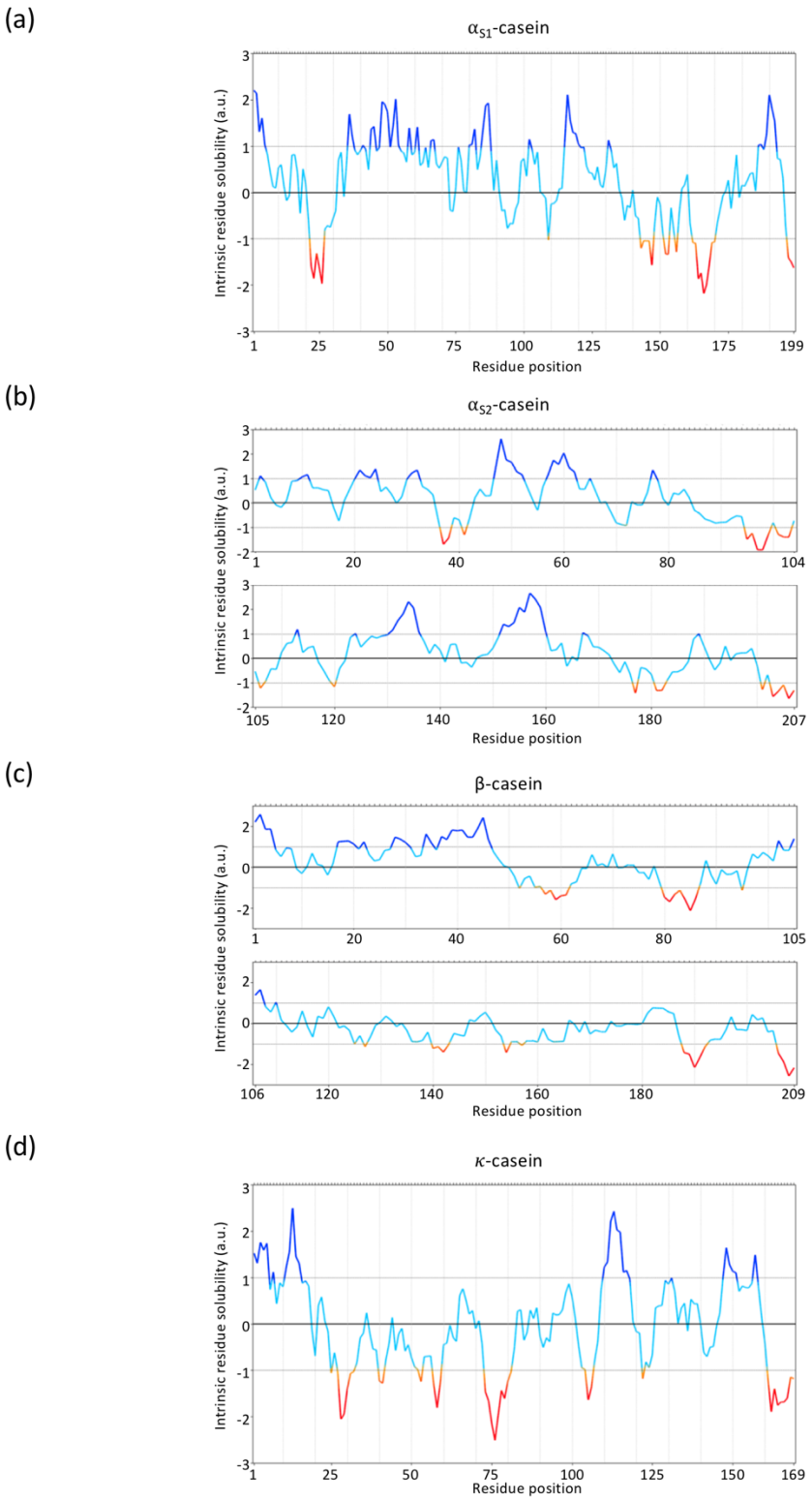

Figure 4. Protein solubility profiles (intrinsic residue solubility versus amino acid sequence) for mature cow caseins as predicted by the CamSol method (https://www-cohsoftware.ch.cam.ac.uk/) [54]. *The regions in red (< -1 a.u. of intrinsic residue solubility)*

correspond to those most likely to aggregate. (a) α_{S1}-casein, (b) α_{S2}-casein, (c) β-casein, (d) κ-casein.

Re-entrant condensation is a phenomenon that describes the overall phase behaviour of biomolecules, polyelectrolytes and colloids arising from charge neutralisation in the presence of multivalent ions [55]. Inside cells, liquid-liquid phase separation is nucleated and seeded by polyanions such as RNA and polyphosphates such as ATP via charge-charge interactions with the proteins concerned. By analogy, in milk, polyanions such as citrate (present at a concentration of 8-10 mM in cow milk) and phosphate may play a similar role in nucleating liquid-liquid phase separation of caseins to form casein micelles. In contrast to the intracellular regime where ATP concentrations are in the mM range, the concentration of ATP is very low in milk as in other extracellular media where ATP is present in the nM range [56]. Hence, ATP is unlikely to have a role in liquid-liquid phase separation of caseins. Consistent with the involvement of anions in casein phase separation, Vollmer *et al.* [57] recently used pentasodium triphosphate (and processing conditions analogous to the production of processed cheese) to disrupt the casein micelle leading to phase separation of the caseins and ultimately conversion of κ-casein (and possibly one or more of the other caseins) into amyloid fibrils.

The binding of caseins to CaP-SLiMs and subsequent nanocluster formation could also create nucleation sites promoting casein liquid-liquid phase separation to form casein micelles. This effect is unique to the extracellular environment as no intracellular CaP nanoclusters have been described.

As mentioned above, α_{S2}-casein and κ-casein form amyloid fibrils under physiological conditions in the absence of the other caseins and CaP. The alteration of Ser to Ala in the CaP-SLiMs of α_{S2}-casein reduced its p_{LLPS} value to 0.18 from 0.45 (unphosphorylated form) and 0.27 (Ser to Glu form) (**Table 1**). The Ser to Ala variant of α_{S2}-casein only has one very short DPR and aggregation hot-spot (M4-E5) in its droplet-promoting propensity profile, compared to six aggregation hot-spots of nine, one, two, seven, one and ten amino acids in length in the unphosphorylated form (**Figures 5a, b**). Examination of κ-casein reveals a less pronounced effect upon comparing the unphosphorylated form to the Ala variant, which is probably associated with its absence of CaP-SLiMs, and only having one phosphorylation site at S149. However, the p_{LLPS} value and the number of aggregation hot-spots are reduced upon mimicking the removal of phosphorylation with the p_{LLPS} values of 0.62 and 0.52, respectively, for the unphosphorylated and Ser to Ala variant (**Table 1**), and five and four aggregation hot-spots encompassing 25 and 14 amino acids in total, respectively (**Figures 5c, d**). Thus, mimicking the loss of charge at the sites associated with CaP binding in α_{S2}-casein leads to a reduction in its propensity to aggregate and form a droplet. As a result of binding to CaP within casein micelles, α_{S2}-casein is stabilised and amyloid fibril formation is minimised. Weaker stabilisation possibly occurs due to interaction between κ-casein and CaP within casein micelles.

Figure 5

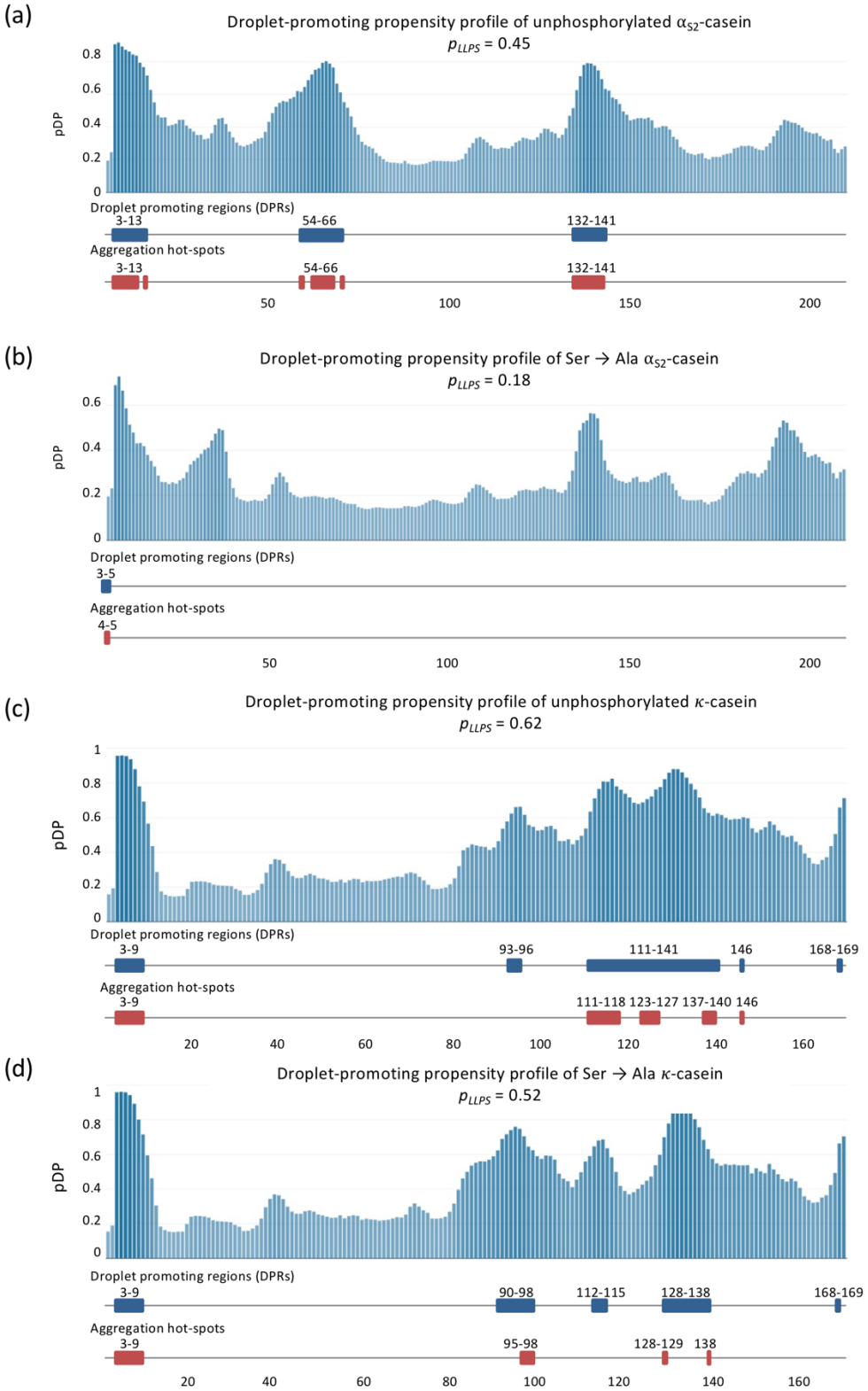

Figure 5. Droplet-promoting propensity profile (pDP values versus amino acid sequence) of unphosphorylated α_{S2}-casein and κ-casein and their Ser to Ala variants. *(a)*

Unphosphorylated α_{S2}-casein. (b) α_{S2}-Casein with 11 Ser to Ala substitutions at positions 8, 9, 10, 16, 56, 57, 58, 61, 129, 131 and 143. (c) Unphosphorylated κ-casein. (d) κ-casein with a Ser to Ala substitution at position 149. The droplet-promoting regions encompass residues above the p_{LLPS} threshold of 0.61. Aggregation hot-spots were identified by FuzDrop based on the likelihood of changing disordered to ordered interactions [33].

Liquid-liquid phase separation, caseins, amyloid fibrils and molecular chaperones

Under physiological conditions, association of caseins into micelles prevents α_{S2}- and κ-casein (and possibly the other caseins) from forming amyloid fibrils due to the molecular chaperone action of the caseins [18, 19, 58]. In addition, casein micelle formation prevents the sequestered nanoclusters of CaP from forming pathological calcified deposits during the storage of milk in the cisterns and ducts of the mammary gland. So successful is this mechanism that a mother may go through repeated cycles of mammary gland development, lactation and tissue remodelling during involution and quiescence so that the gland remains functional throughout her reproductive life [59].

Many intrinsically disordered proteins that undergo phase separation are also prone to amyloid fibril formation either upon aging or under suitable solution conditions. In this case, intrinsically disordered proteins convert from their native to amyloid state via a metastable droplet state (**Figure 6)** [45]. Even when the condensed droplet and amyloid states are thermodynamically more stable than the native state, they are separated by free energy barriers that delay the condensation process. Within the casein micelle, stabilisation between the caseins due to multivalent interactions increases the free energy barrier for converting caseins from the droplet to amyloid states, particularly for α_{S2}- and κ-casein. Thus, these interactions impede the conversion to the amyloid state.

Figure 6

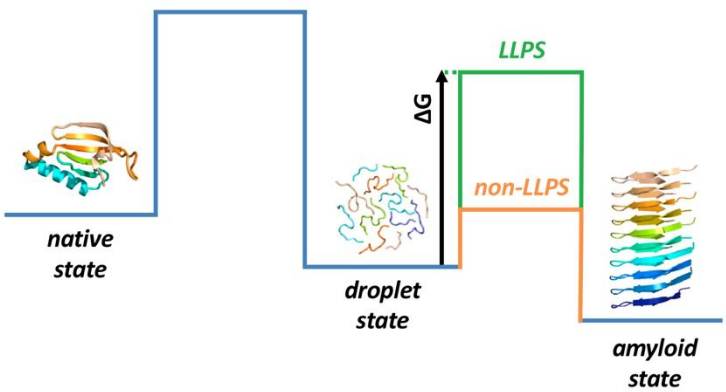

Figure 6. Free energy profile for the conversion of a native protein to its amyloid state via the droplet state. *Above the critical concentration for condensation, the droplet and amyloid states are lower in free energy than the native state. Within a membraneless organelle, the protein is trapped in the droplet state by high free energy barriers (blue and green). In this model, when cow α_{S2}- and κ-casein are incorporated into casein micelles, the free energy*

minimum corresponding to the liquid condensate is stabilised by the many transient, multivalent interactions with the other caseins via a chaperone action that prevents α_{S2}- and κ-casein from converting to the amyloid state. Relatively strong binding of the majority of α_{S1}-, α_{S2}- and β-caseins to the CaP nanoclusters also reduces the possibility of their conversion into amyloid fibrils. In the absence of other caseins, the free energy barrier for conversion to the amyloid state is low (orange), either by a destabilisation of the transition state (as shown) or of the droplet state, and α_{S2}- and κ-casein can progress to the amyloid state. From Hardenberg et al. [45].

The structure and chaperone ability of caseins are analogous in many ways to those of the unrelated, intracellular small heat-shock proteins (sHsps) [8, 10, 60]. Furthermore, some of the human sHsps are involved in phase separation processes [61]: (i) HspB2 undergoes concentration-dependent liquid-liquid phase separation in mammalian cells which is regulated by its partner sHsp, HspB3, with which it forms a stable 3:1 HspB2:HspB3 tetramer *in vitro* [62], (ii) HspB8 plays a central role in maintaining the liquid properties of FUS droplets, and (iii) HspB1 (Hsp27) and HspB5 (αB-crystallin) are recruited into intracellular membraneless organelles upon heat shock, a process which is regulated by their phosphorylation. The p_{LLPS} values of these sHsps reflect their phase separation capabilities (**Table 2**) [45]. Like caseins, sHsps form large, dynamic, heterogeneous and polydisperse aggregates and contain extensive disorder in their N- and C-terminal regions [63] that are proposed to regulate sHsp involvement in phase separation [61].

Via their chaperone action, caseins and sHsps prevent the amorphous and amyloid fibrillar aggregation of their client proteins [60, 64]. Caseins inhibit proteins from aggregating by interacting with them at the early stages of their misfolding pathways, as occurs with sHsps [63, 65, 66]. These interactions may lead to the association between the molecular chaperones and their client proteins. For this reason, caseins and sHsps are sometimes classified as holdase chaperones because they have no ability to refold client proteins, for example via ATP hydrolysis.

Multiple short peptide regions of the major sHsp, HspB5 (αB-crystallin), exhibit chaperone activity on their own, i.e., as isolated, unstructured peptides [67]. The chaperone interactions are both hydrophobic and hydrophilic in nature and are transient and dynamic. During the chaperone action with a client protein, multivalent interactions occur for these regions within whole (intact) HspB5. Intriguingly, many of the peptides involved in chaperone action also form part of the interface involved in subunit-subunit interactions within the HspB5 oligomer [68]. Thus, there is a commonality of regions participating in the multivalent interactions between the subunits of HspB5 in its oligomeric state and the interactions of HspB5 with client proteins during the chaperone action. The multivalent interactions that characterise the association of caseins within the casein micelle are likely also responsible for interacting with destabilised client proteins during the chaperone action of caseins. However, due to the absence of a significant number of hydrophobic and order-promoting residues and HO-SLiMs in caseins, the multivalent interactions involved in chaperone interactions are mainly hydrophilic in nature [6, 8].

Conclusions and future directions

Caseins are intrinsically disordered proteins that associate in milk to form casein micelles via transient multivalent interactions. Small ions and post-translational modifications, in particular phosphorylation that enables multiple casein protein-CaP nanocluster interactions, contribute to these interactions and hence to casein micelle formation. As a result of these properties and behaviour, caseins exhibit marked similarities with intrinsically disordered proteins that undergo intracellular liquid-liquid phase separation to form membraneless organelles. Within casein micelles, multivalent interactions facilitate the chaperone action of caseins to prevent their conversion into amyloid fibrils. Thus, one could consider that caseins are balanced between the native, droplet and amyloid fibrillar states (**Figure 6**).

A schematic summary of the conversion of caseins via phase separation into casein micelles is presented in **Figure 7**. On the basis of the arguments presented here, casein micelles can be described as extracellular condensates. By comparing these condensates with extracellular membrane-bound organelles (e.g. exosomes, shedding microvesicles and apoptotic blebs), one could characterise casein micelles as membraneless organelles formed extracellularly. The existence of extracellular condensates would fill a conceptual gap in biophysical terms because inside the cell proteins can exist in three states (native, liquid-like condensate, and solid-like amyloid) [45], while outside the cell, so far, proteins have been shown to populate only two states (native and amyloid).

The likelihood that caseins undergo liquid-liquid phase separation to form casein micelles via a process that has many similarities to the formation of intracellular biomolecular condensates may provide novel opportunities for researchers investigating casein structure and function and their interactions, and in the utilisation of caseins in the dairy and food industries.

Casein micelles as extracellular liquid-like condensates

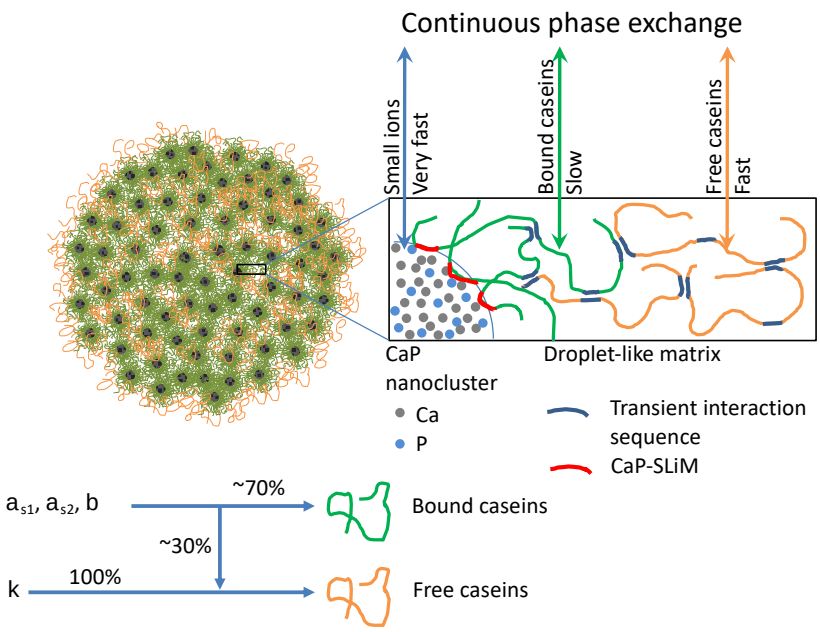

Figure 7. A schematic illustration of casein micelles as extracellular condensates containing intrinsically disordered caseins, CaP nanoclusters and small ions such as Ca^{2+}. *An expanded snapshot of the liquid droplet-like structure near the surface of one of the CaP nanoclusters in the multivalent model of casein micelles [11,14] is also shown. Droplet-promoting regions involved in transient interactions between the caseins, such as those predicted by the FuzDrop method (Figures 3 and 5), are shown in blue. Small ions, caseins bound directly to the CaP surface via CaP-SLiMs (in red) (~70% of α_{S1}, α_{S2} and β) and the remaining free caseins (~30% of α_{S1}, α_{S2} and β and 100% of κ) can exchange with the continuous phase over a range of timescales. According to this model, free casein molecules not bound to CaP nanoclusters near the micelle surface may readily exchange with casein molecules in the serum, while those bound to CaP nanoclusters may form more gel-like assemblies.*

Nutritionally, casein micelles are deconstructed and utilised as a readily accessible and concentrated source of proteins and calcium. Casein micelle formation in milk also protects the mother and her mammary glands from the adverse effects of inadvertent biocalcification and toxic casein amyloid fibril formation [59, 69] due to casein sequestration of CaP nanoclusters and the chaperone action of the caseins. The former may be a specific example of a more general physiological mechanism [12, 70]. In this sense as SCPPs, caseins are involved in the control of biocalcification. SCPPs also include salivary proteins, proteins involved in bone and tooth formation (e.g. amelogenin) and the widely distributed secreted phosphoprotein 1 or osteopontin (OPN). All are known, or predicted to be, intrinsically disordered and several of them form amyloid fibrils. Most have SLiMs that bind CaP when phosphorylated and, experimentally, the N-terminal phosphopeptide of OPN forms CaP nanocluster complexes [71]. Future computational and experimental studies will investigate the interactions between caseins themselves and within casein micelles. Similar studies could be undertaken with other SCPPs to determine their propensity to phase separate to form extracellular condensates.

Author contributions

JAC and CH conceived the topic; AH and MF wrote the software and undertook the calculations; AH, MF and MV provided input on liquid-liquid phase separation and membraneless organelles, JAC and CH on caseins; JAC wrote the draft manuscript which AH, MF, MV and CH revised extensively.

Acknowledgements

MF acknowledges the support of Associazione Italiana per la Ricerca sul Cancro IG 2021 ID 26229.

References

1. Das RK, Ruff KM, Pappu RV (2015). Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* 32, 102-112.

2. Harmon, T.S., Holehouse, A.S., Rosen, M.K. and Pappu, R.V. (2017). Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *eLife* 6, e30294.

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

4. Syme, C.D., Blanch, E.W., Holt, C., Jakes, R., Goedert, M., *et al.* (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.

5. van der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., *et al.* (2014). Classification of intrinsically disordered regions and proteins. *Chem. Rev.* 114, 6589–6631.

6. Carver, J.A. and C. Holt, C. (2020). Functional and dysfunctional folding, association and aggregation of caseins. *Adv. Prot. Chem. Struct. Biol.* 118, 163-216.

7. Sormanni, P., Piovesan, D., Heller, G.T., Bonomi, M., Kukic, P. *et al.* (2017). Simultaneous quantification of protein order and disorder. *Nat. Chem. Biol.* 13, 339-342.

8. Holt, C., Raynes, J.K. and Carver, J.A. (2019). Sequence characteristics responsible for protein-protein interactions in the intrinsically disordered regions of caseins, amelogenins and small heat-shock proteins. *Biopolymers* 110, e23319.

9. Tompa, P. and Fuxreiter, M. (2008). Fuzzy complexes: Polymorphism and structural disorder in protein–protein interactions. *Trends Biochem. Sci*. 33, 2-8.

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

11. Holt, C. and Carver, J.A. (2022). A quantitative multivalent model of the structure, size distribution and composition of the casein micelles of cow milk. *Int. Dairy J*. 126, 105292.

12. Lenton, S., Wang, Q., Nylander, T., Teixeira, S.C.M. and Holt, C. (2020). Structural biology of calcium phosphate nanoclusters sequestered by phosphoproteins. *Crystals* 10, 755.

13. Bijl, E., Huppertz, T., van Valenberg, H. and Holt, C. (2019). A quantitative model of the bovine casein micelle: ion equilibria and calcium phosphate sequestration by individual caseins in bovine milk. *Eur. Biophys. J.* 48, 45-59.

14. Holt, C. (2021). A quantitative calcium phosphate nanocluster model of the casein micelle: the average size, size distribution and surface properties. *Eur. Biophys. J.* 50, 847-866.

15. Knowles, T.P.J., Vendruscolo, M. and Dobson, C.M. (2014). The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.* 15, 384–396.

16. Chiti, F. and Dobson, C.M. (2017). Protein misfolding, amyloid formation, and human disease: A summary of progress over the last decade. *Ann. Rev. Biochem*. 86, 27-68.

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

18. Thorn, D.C., Meehan, S., Sunde, M., Rekas, A., Gras, S.L., *et al.* (2005). Amyloid fibril formation by bovine milk κ-casein and its inhibition by the molecular chaperones α_s- and β-casein. *Biochemistry* 44, 17027-17036.

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

20. Bahraminejad, E. (2019). Ph.D. thesis, The Australian National University.

21. Bhattacharyya, J. and Das, K. P. (1999). Molecular chaperone-like properties of an unfolded protein, αS-casein. *J. Biol. Chem.* 274, 15505-15509.

22. Morgan, P.E., Treweek, T.M., Lindner, R.A., Price, W.E. and Carver, J.A. (2005). Casein proteins as molecular chaperones. *J. Agric. Food Chem.* 53, 2670-2683.

23. Fuxreiter, M. and Vendruscolo, M. (2021). Generic nature of the condensed states of proteins. *Nat. Cell Biol.* 23**,** 587–594.

24. Brangwynne, C.P., Eckmann, C.R., Courson, D.S., Rybarska, A., Hoege, C., *et al.* (2009). Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* 324, 1729-1732.

25. Brangwynne, C.P., Tompa, P. and Pappu, R.V. (2015). Polymer physics of intracellular phase transitions*. Nat. Phys.* 11, 899-904.

26. Banani, S.F., Lee, H.O., Hyman, A.A. and Rosen, M.K. (2014). Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* 18, 285–298 (2017).

27. Gomes, E. and Shorter, J. (2019). The molecular language of membraneless organelles. *J. Biol. Chem*. 294, 7115-7127.

28. You, K., Huang, Q., Yu, C., Shen, B., Sevilla, C., *et al*. (2020). PhaSepDB: a database of liquid-liquid phase separation related proteins. *Nucleic Acids Res.* 48, D354-D359.

29. Parker, R. and Seth, U. (2007). P bodies and the control of mRNA translation and degradation. *Mol. Cell* 25, 635-646.

30. Kedersha, N., Cho, M.R., Li, W., Yacono, P.W., Chen, S., *et al.* (2000). Dynamic shuttling of TIA-1 accompanies the recruitment of mRNA to mammalian stress granules. *J. Cell Biol.* 151, 1257-1268.

31. Buchan, J.S. and Parker, R. (2009). Eukaryotic stress granules: the ins and outs of translation. *Mol. Cell* 36, 932-941.

32. Li, P., Banjade, S., Cheng, H.C., Kim, S., Chen, B., *et al.* (2012). Phase transitions in the assembly of multivalent signalling proteins. *Nature* 483, 336–340.

33. Vendruscolo, M. and Fuxreiter, M. (2021). Sequence determinants of the aggregation of proteins within condensates generated by liquid-liquid phase separation. *J. Mol. Biol*. 434, 167201.

34. Dinkel, H., Van Roey, K., Michael, S., Kumar, M., Uyar, B., *et al.* (2016). ELM 2016-data update and new functionality of the eukaryotic linear motif resource. *Nucl. Acids Res.* 44, D294-D300.

35. Ren, S., Uversky, V.N., Chen, Z., Dunker, A.K. and Obradovic, Z. (2008). Short linear motifs recognized by SH2, SH3 and Ser/Thr kinase domains are conserved in disordered protein regions. *Bmc Genomics* 9, S26.

36. Van Roey, K., Uyar, B., Weatheritt, R. J., Dinkel, H., Seiler, M., *et al.* (2014). Short inear motifs: Ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem. Rev.* 114, 6733-6778.

37. Weatheritt, R.J., Davey, N.E. and Gibson, T.J. (2012). Linear motifs confer functional diversity onto splice variants. *Nucl. Acids Res.* 40, 7123-7131.

38. Delaforge E., Kragelj, J., Tengo, L., Palencia, A., Milles, S., *et al.* (2018). Deciphering the dynamic interaction profile of an intrinsically disordered protein by NMR exchange spectroscopy. *J. Am. Chem. Soc.* 140, 1148-1158.

39. Miskei M., Gregus A., Sharma R., Duro N., Zsolyomi F. and Fuxreiter M. (2017). Fuzziness enables context dependence of protein interactions. *FEBS Lett.* 591, 2682-2695.

40. Liu, D.Z., Weeks, M.G., Dunstan, D.E. and Martin, G.J.O. (2013). Temperature-dependent dynamics of bovine casein micelles in the range 10-40 degrees C. *Food Chem.* 141, 4081-4086.

41. Yang, S., Tyler, A.I.I., Ahrné, L. and Kirkensgaard, J.J.K. (2021). Skimmed milk structural dynamics during high hydrostatic pressure processing from in situ SAXS. *Food Res. Int.* 147, 110527.

42. Weng, J. and Wang, W. (2020). Dynamic multivalent interactions of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* 62, 9-13.

43. Hughes, M.P., Sawaya, M.R., Boyer, D.R., Goldschmidt, L., Rodriguez, J.A., *et al.* (2018). Atomic structures of low-complexity protein segments reveal kinked β sheets that assemble networks. *Science* 359, 698-701.

44. Hughes, M.P., Goldschmidt, L. and Eisenberg, D.S. (2021). Prevalence and species distribution in genomes of the low-complexity, amyloid-like, reversible, kinked segment structural motif in amyloid-like fibrils. *J. Biol. Chem.* 297, 101194.

45. Hardenberg, M., Horvath, A., Ambrus, V., Fuxreiter, M. and Vendruscolo, M. (2020). Widespread occurrence of the droplet state of proteins in the human proteome. *Proc. Natl. Acad. Sci. USA* 117, 33254.

46. Miskei, M., Horvath, A., Vendruscolo, M. and Fuxreiter, M. (2020). Sequence based prediction of fuzzy protein interactions, *J. Mol. Biol.* 432, 2289-2303.

47. Horvath A., Miskei M., Ambrus V., Vendruscolo M. and Fuxreiter M. (2020). Sequence-based prediction of protein binding mode landscapes. *PLoS Comput. Biol.* 16, e1007864.

48. Walsh, I., Martin, A.J.M., Di Domenico, T. and Tosatto, S.C.E. (2012). ESpritz: Accurate and fast prediction of protein disorder. *Bioinformatics* 28, 503- 509.

49. Holt, C. (2016). Casein and casein micelle structures, functions and diversity in 20 species. *Int. Dairy J.* 60, 2-13.

50. Fang, Z. H., Bovenhuis, H., Delacroix-Buchet, A., Miranda, G., Boichard, D., *et al.* (2017). Genetic and nongenetic factors contributing to differences in alpha(s)-casein phosphorylation isoforms and other major milk proteins. *J. Dairy Sci.* 100, 5564-5577.

51. Ecroyd, H., Koudelka, T., Thorn, D.C., Williams, D.M., Devlin, G., *et al.* (2008). Dissociation from the oligomeric state is the rate-determining step in amyloid fibril formation by κ-casein. *J. Biol. Chem.* 283, 9012-9022.

52. Thorn, D.C., Bahraminejad, E., Grosas, A.B., Koudelka, T., Hoffmann, P., *et al.* (2021). Native disulphide-linked dimers facilitate amyloid fibril formation by bovine milk α_{S2}-casein. *Biophys. Chem.* 270, 106530.

53. Tuttle, M.D., Comellas, G., Nieuwkoop, A.J., Covell, D.J., Berthold, D.A., et al. (2016). Solid-state NMR structure of a pathogenic fibril of full-length human alpha-synuclein. Nat. Struct. Mol. Biol. 23, 409-415.

54. Sormanni, P., Aprile, F.A. and Vendruscolo, M. (2015). The CamSol method of rational design of protein mutants with enhanced solubility. *J. Mol. Biol.* 427, 478-490.

55. Zhang, F., Roosen-Runge, F., Sauter, A., Wolf, M., Jacobs, R.M.J. and Schreiber, F. (2014). Reentrant condensation, liquid-lquid phase separation and crystallization in protein solutions induced by multivalent metal ions. *Pure App. Chem*. 86, 191-202.

56. Fuentes, E. and Palomo, I. (2015). Extracellular ATP metabolism on vascular endothelial cells: A pathway with pro-thrombotic and anti-thrombotic molecules. *Vascul. Pharmacol.* 75, 1-6.

57. Vollmer, A.H., Kieferle, I., Pusl, A. and Kulozik, U. (2021). Effect of pentasodium triphosphate concentration on the physicochemical properties, microstructure, and formation of casein fibrils in processed cheese. *J. Dairy Sci.* 104, 11442-11456.

58. Treweek, T.M., Thorn, D.C., Price, W.E. and Carver, J.A. (2011). The chaperone action of bovine milk α_{s1}- and α_{s2}-caseins and their associated form α_s-casein. *Arch. Biochem. Biophys.* 52, 42-52.

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

60. Carver, J.A., Ecroyd, H., Truscott, R.J.W., Thorn, D.C. and Holt, C. (2018). Proteostasis and the regulation of intra and extracellular protein aggregation by ATP-independent molecular chaperones: lens α-crystallins and milk caseins. *Acc. Chem. Res.* 51, 745-752.

61. Carra, S. Alberti, S., Benesch, J.L.P., Boelens, W., Buchner, J., *et al.* (2019). Small heat shock proteins: multifaceted proteins with important implications for life. *Cell Stress Chap.* 24, 295-308.

62. Clark, A.R., Vree Egberts, W.T., Kondrat, F.D., Hilton, G.R., Ray, N.J., *et al*. (2018). Terminal regions confer plasticity to the tetrameric assembly of human HspB2 and HspB3. *J. Mol. Biol.* 430, 3297-3310.

63. Treweek, T.M., Meehan, S., Ecroyd, H. and Carver, J.A. (2015). Small heat-shock proteins: important players in regulating cellular proteostasis. *Cell. Mol. Life Sci.* 72, 429-451.

64. Sanders, H.M., Jovcevski, B., Carver, J.A. and Pukala, T.L. (2020). The molecular chaperone β-casein prevents amorphous and fibrillar aggregation of α-lactalbumin by stabilisation of dynamic disorder. *Biochem. J.* 477, 629-643.

65. Hatters, D.M., Lindner, R.A., Carver, J.A. and Howlett, G.J. (2001). The molecular chaperone, α-crystallin, inhibits amyloid formation by apolipoprotein C-II. *J. Biol. Chem.* 276, 33755-33761.

66. Carver, J.A., Lindner, R.A., Lyon, C. Canet, D., Hernandez, H., *et al.* (2002). The interaction of the molecular chaperone α-crystallin with unfolding α-lactalbumin: a structural and kinetic spectroscopic study. *J. Mol. Biol.* 318, 812-827.

67. Ghosh, J.G., Estrada, M.R. and Clark, J.I. (2005). Interactive domains for chaperone activity in the small heat shock protein, human alphaB crystallin. *Biochemistry* 44, 14854-14869.

68. Ghosh, J.G. and Clark, J.I. (2005). Insights into the domains required for dimerization and assembly of human alphaB crystallin. *Prot. Sci.* 14, 684-695.

69. Dehle, F.C., Ecroyd, H., Musgrave, I.F. and Carver, J.A. (2010). αB-crystallin inhibits the cell toxicity associated with amyloid fibril formation by κ-casein and the amyloid-β peptide. *Cell Stress Chap.* 15, 1013-1026.

70. Holt, C. (2013). Unfolded phosphopolypeptides enable soft and hard tissues to coexist in the same organism with relative ease. *Curr. Opin. Struct. Biol.* 23, 420-425.

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

Table 1. *Droplet-driving probability (p_{LLPS}) of mature cow caseins in their native and most prevalent phosphorylated forms (as determined by phosphomimics, i.e., substituting Glu for Ser at the phosphorylation sites). The substitution of Ala at the main sites of phosphorylation simulates the effect of binding to the CaP nanoclusters. The mature sequences are formed when signal sequences of 15 amino acid residues are cleaved from the N-termini of α_{S1}-, α_{S2}- and β-caseins. In κ-casein, 21 residues are removed. Spontaneous liquid-liquid phase separation is indicated in bold, droplet clients in italics.*

Casein	Unphosphorylated	Phosphomimic (Ser to Glu)	CaP-bound (Ser to Ala)
α_{S1}	**0.90**	**0.91** (8S→E)	**0.82** (8S→A)
α_{S2}	0.45	0.28 (11S→E)	0.18 (11S→A)
β	**0.62**	*0.59* (5S→E)	*0.54* (5S→A)
κ	**0.62**	*0.56* (1S→E)	*0.52* (1S→A)

Table 2. *Droplet-driving probability (p_{LLPS}) of selected human sHsps (Data set S7 in Hardenberg et al. [45]). Spontaneous droplet formation is indicated in bold, clients in italic.*

sHsp	p_{LLPS}
HspB1	**0.64**
HspB2	**0.80**
HspB3	0.25
HspB5	*0.51*
HspB8	**0.96**