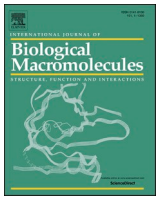
International Journal of Biological Macromolecules 286 (2025) 137790

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: www.elsevier.com/locate/ijbiomac



# Strong electrostatic attraction drives milk heteroprotein complex coacervation

Isabel Vinterbladh [a,*], Rima Hachfi Soussi [b,c], Jan Forsman [a], Said Bouhallab [b], Mikael Lund [a,d]

[a] Division of Computational Chemistry, Lund University, Naturvetarvägen 24, SE-223 62 Lund, Sweden
[b] INRAE, Institut Agro, STLO, 65 Rue de Saint Brieuc, 35042 Rennes, France
[c] Université Paris-Saclay, CNRS, Institut Galien Paris-Saclay, 91400 Orsay, France
[d] LINXS - Institute of advanced Neutron and X-ray Science, Lund University, Scheelevägen 19, 223 70 SE-Lund, Sweden

ARTICLE INFO

*Keywords:*
Coacervates
$\beta$−lactoglobulin
Lactoferrin
Metropolis-Hastings Monte Carlo
Parallel tempering
Heteroprotein complex coacervation
Milk proteins

ABSTRACT

Coacervates of oppositely charged milk proteins are used in functional food development, mainly to encapsulate bioactives. To uncover the driving forces behind coacervates formation, we study the association of lactoferrin and $\beta-$ lactoglobulin at amino-acid level detail, using molecular simulations. Our findings show that inter-protein electrostatic interactions dominate and are, surprisingly, equally divided between an isotropic part, due to monopole-monopole attraction of the oppositely charged proteins, and an anisotropic part due to uneven surface charge distributions. In good agreement with recent experimental association constants, the calculated protein-protein interaction free energy is strongly dependent on pH and salt concentration. In addition to thermodynamics, we also investigate amino acid contacts in microstates of trimeric and pentameric protein complexes, and identify interaction hot-spots that drive heteroprotein complex coacervation process.

## 1. Introduction

Numerous studies on complex coacervation have focused on its applications in food and cosmetics production [1–3]. The process by which two or more types of proteins form coacervates is known as heteroprotein complex coacervation (HPCC) [4,5]. These proteins can be of various types and to mention some; polyion, macroion, disordered protein and DNA [6,7]. In an aqueous solution containing oppositely charged proteins, primarily exhibiting electrostatic interactions, spontaneous liquid-liquid phase separation (LLPS) can occur. This phenomenon creates a polymer-rich dense phase, called coacervates, and a less concentrated phase, named the dilute phase [8].

Among heteroprotein complex coacervates, two bovine milk proteins, $\beta$-lactoglobulin (BLG), a weakly acidic globular protein (pI = 5.2) containing two monomers with a molecular weight of 18.3 kDa (162 amino-acids) and lactoferrin (LF), a strongly basic iron-binding (pI = 8.8) protein with a molecular weight of 83 kDa (689 amino-acids, distributed in two homologous lobes) have been extensively studied owing to their biological and functional properties for food and non-food applications [9].

The heteroprotein complex coacervation of these two proteins can only occur under appropriate solution conditions (pH, stoichiometry, ionic strength and protein concentration) [10]. For this study, a constant pH-value of 5.5 is used, found to be optimal for complex coacervation under experimental conditions (see Table 1). A significant charge difference between the two proteins – LF is strongly basic (approximate charge $+22e$ at pH 5.5) and BLG being weakly acidic (approximate charge $-10e$ at pH 5.5) – facilitates HPCC formation. Previous studies show that LF-BLG complex coacervates can encapsulate bioactives, such as vitamins, and efficiently deliver them for biological uptake [11,12]. An advantage of LF-BLG complexes over non-food carriers is that a bio-carrier poses no threat to the ingesting human or creature [2]. Additionally, due to their high viscosity determined experimentally [13], LF-BLG coacervates can be used as texturing agents to avoid the use of exogenous additives and to help develop protein enriched food products. The viscosity of LF-BLG coacervates is 60 Pa·s at 20 °C. As a comparison, the viscosity of honey is approximately 6 times lower.

With the increasing demand of “clean label” functional food products, the LF-BLG coacervates fulfil these requirements. Thus, studying this HPCC is currently a topic with high interest in the community [14].

LF and BLG only undergo LLPS under certain solution conditions of pH, protein stoichiometry and salt concentration [3,11,15,16]. Different studies report varying intervals for these parameters where the coacervation was observed, and even narrower intervals for optimal

* Corresponding author.
*E-mail addresses:* isabel.vinterbladh@compchem.lu.se (I. Vinterbladh), mikael.lund@compchem.lu.se (M. Lund).

https://doi.org/10.1016/j.ijbiomac.2024.137790
Received 26 August 2024; Received in revised form 6 November 2024; Accepted 15 November 2024
Available online 26 November 2024
0141-8130/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

**Table 1**
Overview of reported conditions where BLG and LF form heteroprotein complex coacervates.

| | Ref. [15] | Ref. [16] | Ref. [11] |
|---|---|---|---|
| pH | 5.7–6.2 | 5.5–7.3 | 5.5–6 |
| maximum salt concentration | 20 mM | 100 mM | 25 mM |
| BLG-monomers/LF molar ratio | 4:1 | 3:1 | between 4:1 and 8:1 |

parameters where the coacervation's yield is at its maximum.

An overview with results from previous studies is presented in Table 1.

In this study, we aim to unravel the interaction mechanisms between BLG and LF at a molecular level, providing insights valuable for future experimental investigations and food product development. For this work, Metropolis-Hastings Monte Carlo (MC) simulations were used. By employing MC simulations, we can study the association between single or multiple units of each protein, and thus observe how BLG and LF form trimers and larger clusters. From the simulations, we calculate the interaction free energy of the system and find the free energy minimum. At this protein-protein separation, the interaction hotspots on the protein surfaces for both a trimer and pentamer are identified. Our results provide information on (1) the attraction between the two proteins, (2) specific amino acid to amino acid contacts, and (3) underscores the feasibility of pentamer formation, adding depth to our understanding of protein-protein interactions in complex biological systems.

## 2. Methods

### 2.1. Molecular simulations

We explore the interaction between a BLG dimer and LF using Metropolis-Hastings Monte Carlo (MC) simulations [17,18]. The open source MC framework Faunus is used for the simulations [19], and we employ a protein-protein interaction model that has previously been tested on milk and other proteins [20–22]. An electronic notebook for reproducing the simulation setup is provided as supporting material.

The two protein structures (PDB 1BEB [23] and 1BLF [24]) are kept rigid and coarse grained, such that each amino acid is represented by a single bead of size $\sigma_i$. These beads interact with a combined Lennard-Jones and electrostatic pair potential to account for short-range repulsion and attraction, as well as electrostatic interactions in an aqueous electrolyte solution. The ionization state of each amino acid is fluctuating according to its p$K_a$-value and solution pH; this is done using a constant pH ensemble [25–27]. For all simulations, pH 5.5 was used, matching available experimental data later used for comparison [8]. The total MC system energy due to intermolecular interactions is:

$$U = \sum_{i}^{N-1}\sum_{j=i+1}^{N}\left\{4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + k_BT\frac{\lambda_B z_i z_j}{r_{ij}}e^{-\kappa r_{ij}}\right\} + \mathscr{B}(R) \qquad (1)$$

where $\lambda_B = 7$ Å is the Bjerrum length, $\kappa$ is the inverse screening length, $z$ are residue valencies, $r$ the inter-residue separation, $\varepsilon = 0.05\,k_BT$ the Lennard-Jones interaction strength. This value has been used successfully in other similar studies, providing satisfactory results [21]. Additionally, a uniform dielectric constant is utilized, previous studies have shown that for a coarse-grained system it is a qualified approximation [28–31]. The last term is a bias potential, $\mathscr{B}(R)$, explained below.

The mass centres of the two rotating proteins are fixed on a line, along which they are allowed to translate. In addition to rotational and translation MC moves, move attempts are also performed to (de-)protonate acidic and basic residues [25,32,33]. During simulation, we sample the radial distribution function, $g(R)$, which is related to the angularly averaged potential of mean force, $w(R) = -k_BTlng(R) + w_0$, where the reference state $w_0$ is chosen such that $w(R)\to 0$ for large $R$. This simulation structure is illustrated in Fig. 1.

The following sections further explain the simulation set-up, and are for the technically interested reader. Others may skip to the Results and Discussion Section.

#### 2.1.1. Biased sampling to cover large mass center separation

The proteins, BLG and LF, are oppositely charged and at low salt concentrations, $c_s$, they *strongly associate* due to electrostatic attraction. Because of this, MC predominantly samples short separations and almost no sampling is done at large $R$. This makes it very difficult to determine $w_0$ within reasonable simulation times. As an example, if the interaction free energy is $-16\,k_BT$ when the proteins are in contact, then the relative probability of observing them at long separations is roughly $e^{-16} \sim 10^{-7}$. To promote sampling at large separations, we add a temporary *bias potential* to the MC energy function:

$$\mathscr{B}(R) = -k_BT\lambda_B Z_1 Z_2\left(\frac{sinh(\kappa a)}{\kappa a}\right)^2\left[\frac{1+\kappa a coth(\kappa a)}{R} - \frac{\kappa}{2}\right]e^{-\kappa R} \qquad (2)$$

The observed potential of mean force, $w'(R)$, is now *biased* and hence unphysical, but since $\mathscr{B}(R)$ is isotropic and added along the mass-center separation, $R$, we can simply unbias by subtraction to recover the true and physically sound free energy, $w(R) = w'(R) - \mathscr{B}(R)$. The choice of

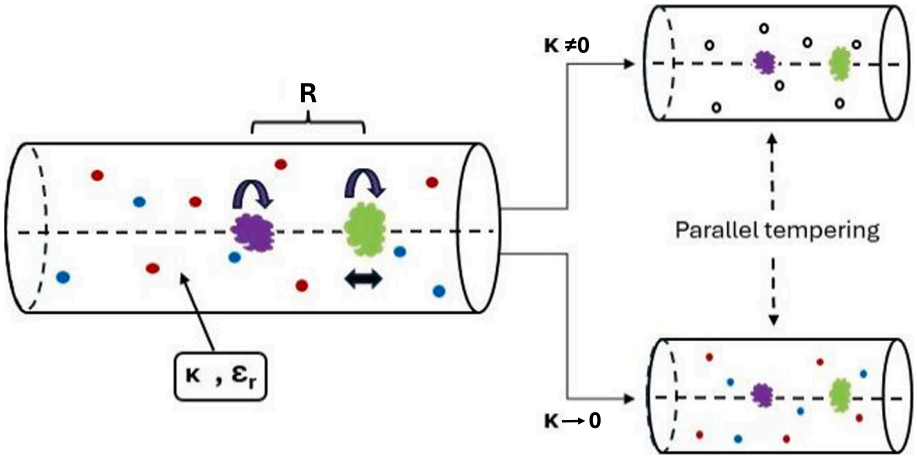



**Fig. 1.** Illustration of the MC simulation method. Proteins (purple and green) are placed on the cylinder's $z$-axis and allowed to rotate. The green off-centre protein can also translate along $z$ and simulations, with different mass centre distances, $R$ are used. Red and blue dots symbolize (optional) explicit salt particles. In parallel tempering simulations, salt particle charges are scaled based on the inverse screening length, $\kappa^{-1}$ so that we effectively sample between fully explicit and fully implicit salt. See text for details. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$\mathscr{B}(R)$ is arbitrary, but most efficient if $w'(R)$ is as flat as possible, whereby all distances become accessible. The above function is the monopole-monopole interaction in an electrolyte solution from Sogame-Ise theory [34,35] and effectively removes a large part of the electrostatic attraction (note the negative sign). $Z_1$ and $Z_2$ denote the average protein valencies and $a = 36$ Å is an effective diameter, chosen such that $w'(R)$ is reasonably flat over a large range of ionic strengths.

To enhance sampling with implicit salt, a parallel tempering move [36] is included, where configurations are interchanged between systems with different salt concentrations. This was used for the implicit salt simulations together with the bias to promote sampling at all centre of mass distances.

#### 2.1.2. Sampling of the electrolyte medium

For *implicit* salt simulation, i.e. $\kappa \neq 0$, the above bias potential is sufficient to obtain converged results for e.g. $w(R)$. However, explicit ions resemble reality better by avoiding the Poisson-Boltzmann approximation used with implicit salt. To confirm the reliability of the implicit ion simulations, grand canonical explicit ion Monte Carlo simulations were performed for a limited number of salt concentrations.

The sampling with explicit ions is executed the same way as the implicit ion case explained above, with the exception that the proteins are kept fixed at $R$ and virtual translation moves [37] along the $z$-axis are used to measure the inter-protein force. These values are numerically integrated to provide the potential of mean force. Ions are unrestricted and can undergo translation MC moves.

#### 2.1.3. Hybrid salt

Due to the irregular protein shape and very strong attraction, the grand canonical explicit salt simulation described above was unable to converge at smaller $R$. When the two proteins are placed close to each other, protein moves are likely rejected due to protein or ion overlap. The explicit ions accumulate close to the proteins to counteract the high electrostatic attraction, where they efficiently block diffusion.

Therefore, in order to sample at these offsets, a parallel tempering simulation was setup. The states represent declining degrees of explicit to implicit solvent. The first state being completely an explicit environment and the last purely an implicit environment. A parameter $\lambda$ with a value of $[0,1]$ was defined. This parameter scales ionic charge, $(1-\lambda)q$, as well as the parameter kappa, $\kappa = \lambda\kappa$, in order to produce the varying explicit to implicit states.

With the help of parallel tempering the states were interchanged during the simulation, making it possible for configurations sampled at the implicit end, where the proteins can rotate more freely, to reach the explicit state and advance the virtual translate sampling.

This system was used for mass-center separations of 50–65 Å. The results were then combined with the output from the simulations with larger separations. From the virtual translation moves the force is obtained and with numerical integration the potential of mean force is calculated.

Moreover, with the proteins in a minimum configuration, a simulation to provide a contact map was performed. Meaning, extracting which amino acids in respective BLG and LF that interact with each other. This is done both for a BLG-LF trimer, and a BLG-LF trimer interacting with another BLG dimer.

## 3. Results and discussion

Using molecular simulations, we investigate how the interaction free energy between BLG and LF varies with changes in salt concentration and protein-protein distance. This analysis provides insights into the strength of interaction between the two proteins under different solution conditions and correlates with experimental association constants. The objective is to gain a deeper understanding of the physical principles underlying BLG and LF complex formation and how solution conditions modular the stability.

### 3.1. Protein-protein interaction free energy

We first use Metropolis-Hastings Monte Carlo simulations to investigate the angular averaged interaction free energy, $w(R)$, as a function of protein mass centre separation, $R$. This is also known as the potential of mean force. Fig. 2a presents simulation results from implicit ion as well as from explicit ion sampling. As expected, salt has a large effect on the protein-protein interaction: at low salt, there is a strong attraction of up to $-17\ k_BT$ respective $-15\ k_BT$. This vanishes at high salt concentrations, which the implicit salt samples at 100 mM shows, in agreement with the experimental observation that coacervates exist only at very low salt concentrations.

Using explicit salt increases the difficulty to sample at high salt concentrations due to a higher density of salt particles, which means that the likelihood of a salt particle blocking the sampling of protein movements increase. Thus, only lower salt concentrations could be sampled. Since the coacervates only exist at low salt concentration, this is a minor concern. Additionally, a reason for performing the grand canonical explicit salt simulations were to affirm the results from the implicit salt sampling, by confirming with a more elaborate approach.

The salt particles, in the explicit ions simulations, screen more efficiently the electrostatic attraction of the proteins from each other, whereby a weaker interaction free energy results. However, the energies are still in the same regime and follow its implicit counterpart well, which indicates that the implicit simulations can be seen as a good approximation.

Recent ITC experiments report on the salt dependent association constant between BLG and LF [8] and, in qualitative agreement with our simulations, the association decreases with increasing salt. Fig. 2b illustrates a *quantitative* comparison of salt dependence in the ITC experiments and our simulations. Experimentally, salt concentrations smaller than 10 mM were utilized, however computer simulations of the proteins could not be executed with this low salt concentrations. For lower salt concentrations, less salt can screen the proteins from each other leading to very strong attractive forces, producing very large absolute free energies which the simulation is unable to sample. Thus, simulations have not been performed for lower concentrations.

The association constant, $K_a$, can in principle be obtained by a volume integration of the interaction free energy, $w(R)$ [38], but this is numerically difficult since the error grows quadratically with $R$. Instead, we approximate the binding free energy as $w(R_{min}) = min(w(R))$. Both measured and simulated data can then be reported relative to a reference ($^*$) salt concentration set to $c_s^* = 15$ mM. We thus report on

$$\Delta\Delta G_{exp} = \Delta G_{exp} - \Delta G_{exp}^* = -k_BT \ln \frac{K_a}{K_a^*} \tag{3}$$

and

$$\Delta\Delta G_{sim} \approx w(R_{min}) - w^*(R_{min}) \tag{4}$$

where $R_{min} = 56$ Å.

Data points in the overlapping salt concentration region fits reasonably well and show a clear trend of decreasing attraction as the salt concentration is increased. At high salt, the simulations show a plateau, which indicates that electrostatic interactions are fully screened. Note that this result is obtained with mean-field electrostatics, and a less approximate theory may modulate this observation slightly [20].

We confirm that an implicit salt model using the Poisson-Boltzmann approximation, aligns with results from explicit salt simulations, as seen in Fig. 2a. While the latter is much less approximate, it is also much harder to sample, since counter-ions accumulate around the highly charged proteins. This hampers protein translation and rotation during simulation, resulting in poor sampling at smaller separations. A significant challenge throughout this study, was to sample extreme inter-protein attractions at low salt concentrations. This led to the

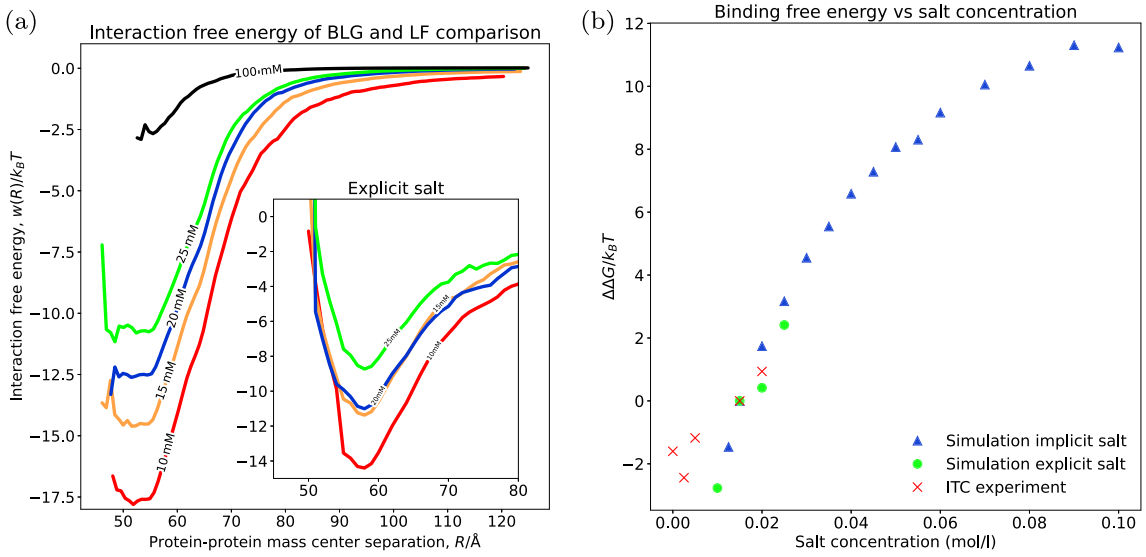



**Fig. 2.** The first plot shows the interaction free energy, whereas the second one displays relative energies where the reference salt concentration of 15 mM is subtracted from the interaction energies. Each value represents the relative binding free energy at the protein-protein separation 56 Å for a specific salt concentration. (a) Simulations of protein-protein interactions to study the interaction free energy of BLG and LF in both implicit and explicit salt. The mean standard deviation is 0.117 and 0.126 for implicit respective explicit salt, see SI for details. (b) Experimental ITC results [8] are used for comparison, where the relative binding free energies are plotted.

development of a parallel tempering method for sampling smaller protein-protein mass separations in the explicit ion environment. The closer the proteins are placed, the higher the attraction is and the harder it gets to find accepted states. This constrained us to only study salt concentrations of 10 mM and higher. Simulations at lower concentrations were tested, however moves and sampling were rarely accepted in these systems, due to the high attractions, providing unreliable and insufficient data. In order to conquer this problem, incorporating Molecular Dynamics (MD) with the MC simulation could perhaps help with sampling extreme interactions and make it possible to obtain simulation results at lower salt concentrations. This is something to be considered in prospective projects.

### 3.2. Charge anisotropy

In this section, we analyse the electrostatic energy between the two proteins in terms of an *electrostatic multipole expansion* [39–41]. First, consider a fixed mass-center separation, $R$. During simulation, the two proteins rotate, and we can sample the direct inter-protein Coulomb interaction,

$$U_{exact}(R) = k_B T \left\langle \sum_i^{N_{LF}} \sum_j^{N_{\beta Lac}} \lambda_B z_i z_j \Big/ r_{ij} \right\rangle \tag{5}$$

For each simulated configuration, still at a fixed $R$, we calculate the instantaneous protein monopole; dipole; and quadrupolar moments. Using a multipole expansion, we use these moments to *approximate* the electrostatic energy and the sum of all terms (ion-ion, ion-dipole etc.) should ideally sum up to $U_{exact}$, in the limit of infinitely many terms and at long separations [42,43]. From the simulations, we have access to all $R$, but here limit the analysis to the free energy minimum which is roughly at $R = 56$ Å. For more details about the analysis, please refer to [39].

Fig. 3 illustrates the different terms as a percentage of the *exact* energy. Unsurprisingly, the attractive ion-ion term gives the single largest contribution, while the ion-dipole and ion-quadrupole terms give roughly equal contributions and are both attractive. Importantly, this analysis shows that anisotropic contributions are responsible for more than half of the total electrostatic interaction. Hence, the two proteins

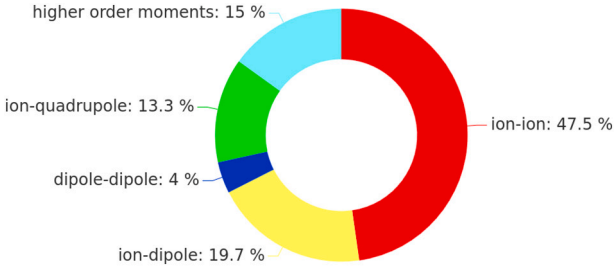



**Fig. 3.** Doughnut chart representation of the relative multipolar contributions obtained from MC simulations at 10 mM salt. Percentages are with respect to the exact electrostatic energy, Eq. (5). The full analysis for this model is presented in SI as a (Jupyter) notebook.

are subject to significant alignment and that these to a large part are due to ion-dipole and ion-quadrupole interactions. Everything not captured by the multipole expansion is assigned to *higher order moments*.

By summing all multipolar terms, the total gets quite close to the exact electrostatic energy (Eq. (5)) despite that we have settled on a short separation (see SI Fig. 3).Thus, multipoles are a good first approximation to mathematically describe charged patches on the protein surface [42,44–46].

### 3.3. Residue-residue contacts

We now investigate the BLG-LF complex when the two proteins are at contact and close to their free energy minimum. This is initially done by defining "contact" when mass centre separation between two amino acids are below 20 Å. Note that in the coarse grained protein model, amino acid diameters are 6–8 Å and that contacts are sampled from implicit salt configurations allowed to rotate in a small interval around the free energy minimum. Possible contacts are calculated within a trimer, between one BLG dimer and one LF, (Fig. 4) as well as between one BLG dimer and one BLG-LF trimer (Fig. 5).

The results are presented as a contact map, as well as by coloured coarse-grained proteins and their secondary structure. Yellow amino acids represents less to no contacts and blue amino acids representing

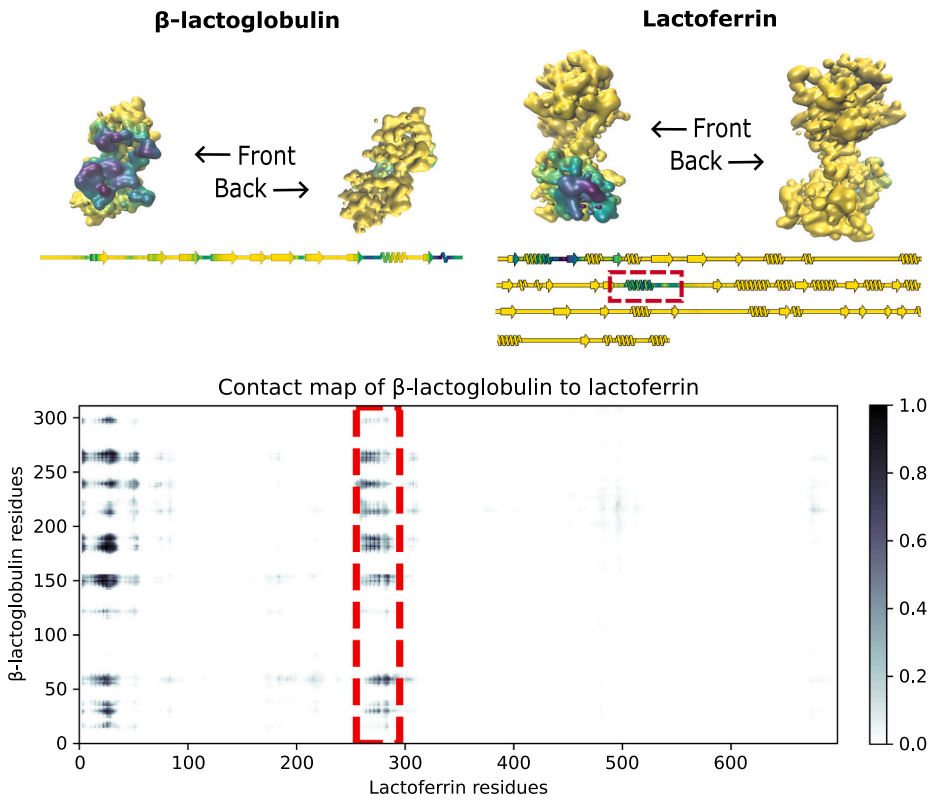



**Fig. 4.** Residue-residue contacts illustrated by protein structure models and contact map. The scale shows the relative number of contacts, darker meaning more. The red rectangle in the secondary structure corresponds to the same marking in the contact map. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

many to maximum amount of contacts. In the secondary structure diagrams, the respective areas of high contact corresponds to the results in the contact maps above. Moreover, comparing the secondary structure diagram to its coarse-grained protein model, we can distinguish which parts of the protein that represent the contact area, both in folded form and in the secondary structure chain. Whereas, the contact maps shows which combination of amino acids that are part of the contact area. The contact maps are generated by averaging the number of observed contacts in the free energy minimum, at $R = 56$ Å. The results represent thermal averages over microstates from a MC simulation at 10 mM salt with a fixed protein-protein mass centre separation (the free energy minimum).

In Fig. 4:bottom, we note two distinct (vertical) bands on LF, centred around residues 20 (HARG) and 285 (PHE). In contrast, BLG, shows many more (horizontal) bands, suggesting that many poses can bind to the same fixed binding sites on LF. Fig. 5:bottom presents similar results. The BLG displays many horizontal lines, illustrating that many different amino acids can be part of the contact. On the trimer, the BLG seems to bind to residue 20 and 285 of the LF, just as in Fig. 4:top. Additionally, the contact area to the BLG dimer in the BLG-LF trimer, is more versatile and a faint grid pattern of amino acids in contact can be distinguished.

### 3.4. Voronoi tessellation

We have used coarse grained MC simulations to find the free energy minimum and thus the contact area of the two studied proteins. Defining the contact area can also be accomplished by using a Voronoi tessellation method [47]. Voronoi tessellation is a partition of a plane into regions called Voronoi cells constructured by points, or "nodes". This tessellation can also be done as a volume partition. Construction of the Voronoi cells can be accomplished by brute-force algorithms, but many more efficient algorithms and methods have been developed [48,49].

Voronota [47], an open-source program, calculates and construct the contacts between and within macro-molecules with the help of Voronoi tessellation. This program was used for comparing and confirming the reliability of our calculated contacts. Voronota needs an atomistic pdb-file as input, in order to calculate and visualise the contact area.

Since our simulations are coarse-grained, we back-converted the coarse-grained structure of the two proteins in their free energy minimum to a full atomistic complex. The visualisation program VMD [50] was used for aligning the coarse-grained structures with an all-atom structure, creating an atomistic copy. With this approach, the side chains, which are absent in the coarse-grained version, cannot be aligned. In order to optimize the protein side chain packing, the online rotamer library program FASPR [51] was used.

The atomistic structure neglect possible protonation changes of the coarse-grained amino acids, resulting in that the same amino acids in the atomistic structure might not represent the same state that the amino acid had in the coarse-grained one. This is a source of error that should be considered in the following analysis.

Contact in our simulation is based on if two amino acids are within a distance threshold, then the contact value of that amino acid is increased. This is done over a trajectory of free energy minimum configurations allowed to rotate. Finally, the contact values are normalised and plotted in a contact map (see Fig. 4).

For our structure, Voronota visualised a contact area between the proteins and listed 58 amino acids, which created 87 inter-protein contact pairs. Fig. 6a illustrates the contact area between the atomistic protein structures with side chains in their free energy minimum. In Fig. 6b a zoom-in of the contact area is presented. Each face of the contact area represents a contact and the colouring illustrate the stress on respective contact point, where red is high and blue low. Voronota does not take pH and temperature into account and with the uncertainties introduced when transforming the coarse-grained structure

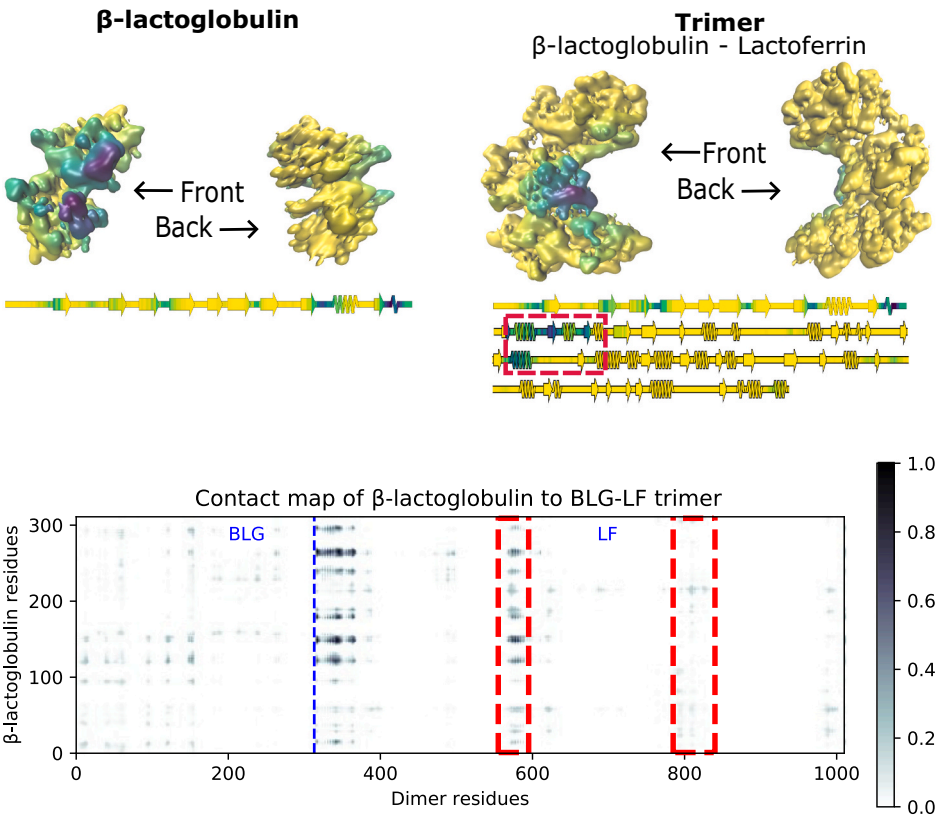



**Fig. 5.** Residue-residue contacts illustrated by protein structure models and contact map. In the contact map, the blue striped line illustrates the last BLG residue and the start of LF in the trimer. The red rectangle in the sequence corresponds to the two red rectangles in the contact map. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

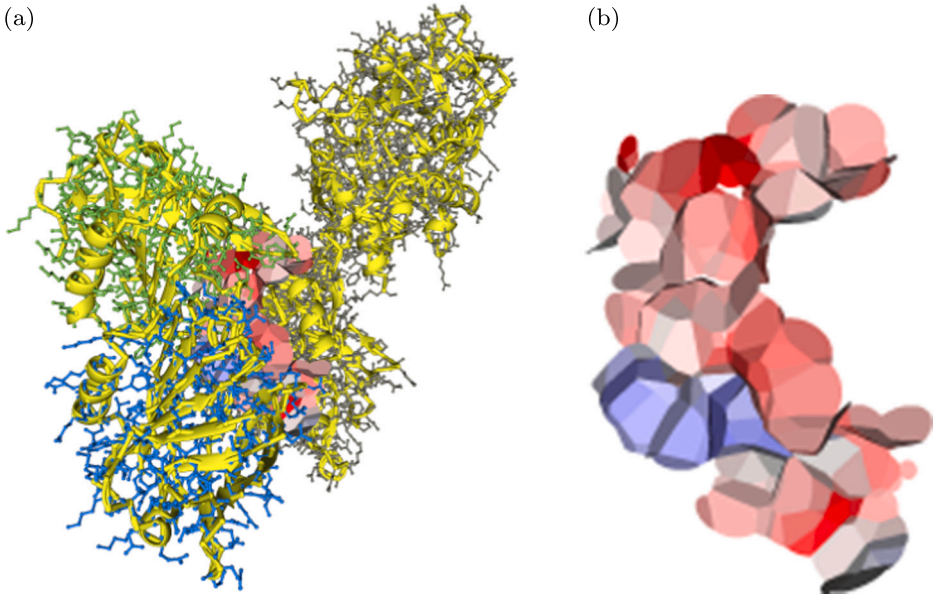



**Fig. 6.** Voronoi tessellation of inter-protein contacts in the LF-BLG complex as obtained using the Voronota software [47].Each face of the surface represents a contact between an atom in each protein. The size of the face shows the contact strength, and the colour, the strain on the contact.

to an atomistic one, we can only assume that with fewer approximations the result would have been more reliable.

Approximately 10 % of the listed contacts by Voronota are not found by our simulation. Our simulation only provides a value of how likely there is a contact between two amino acids. Thus, depending on what value is set as a minimum contact value, more or less amino acids will be listed. Depending on the chosen cut-off value, this differs between 10 to 20 amino acids. However, they are all found to be either adjacent to other defined contacts or to the missing Voronota ones. This makes it hard to discern what cut-off value to use to define an optimal contact area, but even with a somewhat larger contact area it is still comparatively similar to the one defined by Voronota.

## 4. Conclusion

We have investigated the interaction between the milk proteins BLG and LF using molecular MC simulations. Our results show that the association is strongly salt dependent and that the anisotropic and oppositely charged proteins lead to a high percentage of the attraction coming from patchy charge interactions that tend to align the proteins in specific orientations. Our results are in good agreement with experimental data that suggest coacervates formation only at low ionic strengths. Collecting microstates around the free energy minimum provided insight to which amino acids that are involved in the interaction of LF-BLG trimers and BLG-(BLG-LF) pentamers. In both cases, the interaction leads to highly oriented complexes, and we observe distinct interaction hotspots on the protein surfaces. The analysis of residue-residue contacts is based on both contact maps and Voronoi tessellation of contact areas.

Our simulation results corroborate, reinforce and complement at amino acid scale the experimental findings from at least two groups suggesting the existence of a pentamer as a primary unit driving the coacervation process in BLG/LF mixture [5,8,10]. Based on a diversity of complementary and powerful techniques, these concomitant studies suggest the formation of the pentamer through the binding of two BLG dimers on two different sites of one LF, one of higher and one of lower affinity.

Our model incorporates different approximations, and builds on that the studied system mainly exhibits electrostatic interactions. This certainly poses limits, however we have taken these into account by building our model on values and methods from previous studies for similar molecules. Another limitation is that we decide how many proteins to simulate, in a real solution more or less proteins could participate in the coacervation. However, since it is believed that a pentamer is the primary unit, our results can at least show how a single coacervate unit is build up.

In future studies, the optimal molar ratio between the two milk proteins in the coacervates can be investigated. As seen in Table 1, different experimental studies report varying ratios. A Monte Carlo grand canonical simulation or Gibbs ensemble [52] could be used for determining the optimal molar ratio of the proteins. Gibbs ensemble can study phase separation and coexistence, and a grand canonical ensemble can find the thermodynamic equilibrium of a system. By developing these models, we hope to in future studies provide an answer of which molar ratios that are realistically feasible and can create coacervates.

In conclusion, our results will enhance the possibility of further development and research around heteroprotein coacervation, of BLG-LF and other similar proteins, for both computational and experimental studies. Conceivably, these results can promote the development of new functional food products, cosmetics, and food texturing using heteroprotein coacervates.

An electronic (Jupyter) notebook is provided to reproduce the presented simulations and illustrations: https://doi.org/10.5281/zenodo.14223026, in addition to a SI document. Supplementary data to this article can be found online at doi: https://doi.org/10.1016/j.ijbiomac.2024.137790.

## CRediT authorship contribution statement

**Isabel Vinterbladh:** Writing – original draft, Visualization, Validation, Software, Investigation, Formal analysis, Writing – review & editing. **Rima Hachfi Soussi:** Investigation. **Jan Forsman:** Supervision, Methodology. **Said Bouhallab:** Supervision, Conceptualization. **Mikael Lund:** Supervision, Project administration, Investigation, Conceptualization.

## Funding

For financial support, M.L. thanks the Swedish Research Council, grant number 2022-04251. J.F thanks the Swedish Research Council, grant number 2021-04041, for financial support. We gratefully acknowledge the financial support provided by the Regional council of Brittany France and INRAE.

R.H.S. thanks L'Ecole Des Docteurs Bretagne-Loire for its financial contribution.

## Declaration of competing interest

The authors declare there are no competing financial or personal interests.

### *Acknowledgement*

We gratefully thank Kliment Olechnovic for the help with understanding and using the Voronota and VoroContact software packages. For computational resources, we thank LUNARC in Lund.

## Data availability

The data can be reproduces using the Jupyter notebooks available on Github https://doi.org/10.5281/zenodo.14223026.

## References

[1] G.M. Tavares, T. Croguennec, A.F. Carvalho, S. Bouhallab, Milk proteins as encapsulation devices and delivery vehicles: applications and trends, Trends Food Sci. Technol. 37 (1) (2014) 5–20, https://doi.org/10.1016/j.tifs.2014.02.008. https://www.sciencedirect.com/science/article/pii/S0924224414000454.

[2] A.-L. Chapeau, G.M. Tavares, P. Hamon, T. Croguennec, D. Poncelet, S. Bouhallab, Spontaneous co-assembly of lactoferrin and $\beta-$ lactoglobulin as a promising biocarrier for vitamin b9, Food Hydrocoll. 57 (2016) 280–290, https://doi.org/10.1016/j.foodhyd.2016.02.003. https://www.sciencedirect.com/science/article/pii/S0268005X16300339.

[3] C.S. Saraiva, J.S. dos Reis Coimbra, A.V.N. de Carvalho Teixeira, E.B. de Oliveira, R.F. Teóflo, A.R. da Costa, Éverton de Almeida, Alves Barbosa, Formation and characterization of supramolecular structures of $\beta-$ lactoglobulin and lactoferrin proteins, Food Res. Int. 100 (2017) 674–681, https://doi.org/10.1016/j.foodres.2017.07.065. https://www.sciencedirect.com/science/article/pii/S0963996917304040.

[4] E. Kizilay, D. Seeman, Y. Yan, X. Du, P.L. Dubin, L. Donato-Capel, L. Bovetto, C. Schmitt, Structure of bovine $\beta-$ lactoglobulin–lactoferrin coacervates, Soft Matter 10 (2014) 7262–7268, https://doi.org/10.1039/C4SM01333F.

[5] S.E. Flanagan, A.J. Malanowski, E. Kizilay, D. Seeman, P.L. Dubin, L. Donato-Capel, L. Bovetto, C. Schmitt, Complex equilibria, speciation, and heteroprotein coacervation of lactoferrin and $\beta-$ lactoglobulin, Langmuir 31 (5) (2015) 1776–1783, https://doi.org/10.1021/la504020e.

[6] S.J. de Carvalho, R. Metzler, A.G. Cherstvy, Critical adsorption of polyelectrolytes onto charged janus nanospheres, Phys. Chem. Chem. Phys. 16 (2014) 15539–15550, https://doi.org/10.1039/C4CP02207F.

[7] N. Pawar, H. B. Bohidar, Statistical thermodynamics of liquid-liquid phase separation in ternary systems during complex coacervation, Phys. Rev. E 82 (2010) 036107. doi:https://doi.org/10.1103/PhysRevE.82.036107.

[8] R. Soussi Hachfi, P. Hamon, F. Rousseau, M.-H. Famelart, S. Bouhallab, Ionic strength dependence of the complex coacervation between lactoferrin and beta;-lactoglobulin, Foods 12 (5) (2023) 1040, https://doi.org/10.3390/foods12051040. https://www.mdpi.com/2304-8158/12/5/1040.

[9] T. Croguennec, G.M. Tavares, S. Bouhallab, Heteroprotein complex coacervation: A generic process, Adv. Colloid Interf. Sci. 239 (2017) 115–126, complex Coacervation: Principles and Applications, https://doi.org/10.1016/j.cis.2016.06.009, https://www.sciencedirect.com/science/article/pii/S0001868616301452.

[10] P.D.S. Peixoto, G.M. Tavares, T. Croguennec, A. Nicolas, P. Hamon, C. Roiland, S. Bouhallab, Structure and dynamics of heteroprotein coacervates, Langmuir 32 (31) (2016) 7821–7828, https://doi.org/10.1021/acs.langmuir.6b01015.

[11] G.M. Tavares, T. Croguennec, P. Hamon, A.F. Carvalho, S. Bouhallab, Selective coacervation between lactoferrin and the two isoforms of $\beta-$ lactoglobulin, Food Hydrocoll. 48 (2015) 238–247, https://doi.org/10.1016/j.foodhyd.2015.02.027. https://www.sciencedirect.com/science/article/pii/S0268005X15000880.

[12] A.-L. Chapeau, N. Bertrand, V. Briard-Bion, P. Hamon, D. Poncelet, S. Bouhallab, Coacervates of whey proteins to protect and improve the oral delivery of a bioactive molecule, J. Funct. Foods 38 (2017) 197–204, https://doi.org/10.1016/j.jff.2017.09.009.

[13] R. Soussi Hachfi, M.-H. Famelart, F. Rousseau, P. Hamon, S. Bouhallab, Rheological characterization of $\beta-$ lactoglobulin/lactoferrin complex coacervates, LWT 163 (2022) 113577, https://doi.org/10.1016/j.lwt.2022.113577. https://www.sciencedirect.com/science/article/pii/S0023643822005126.

[14] G.M. Tavares, T. Croguennec, P. Hamon, A.F. Carvalho, S. Bouhallab, How the presence of a small molecule affects the complex coacervation between lactoferrin

and β − lactoglobulin, Int. J. Biol. Macromol. 102 (2017) 192–199, https://doi.org/10.1016/j.ijbiomac.2017.04.007. https://www.sciencedirect.com/science/article/pii/S0141813016320773.
[15] Y. Yan, E. Kizilay, D. Seeman, S. Flanagan, P.L. Dubin, L. Bovetto, L. Donato, C. Schmitt, Heteroprotein complex coacervation: bovine β − lactoglobulin and lactoferrin, Langmuir 29 (50) (2013) 15614–15623, https://doi.org/10.1021/la4027464.
[16] S.G. Anema, C.K. de Kruif, Complex coacervates of lactotransferrin and β − lactoglobulin, J. Colloid Interface Sci. 430 (2014) 214–220, https://doi.org/10.1016/j.jcis.2014.05.036. https://www.sciencedirect.com/science/article/pii/S0021979714003427.
[17] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys. 21 (6) (1953) 1087–1092. arXiv:https://pubs.aip.org/aip/jcp/article-pdf/21/6/1087/18802390/1087_1_online.pdf, doi:https://doi.org/10.1063/1.1699114.
[18] W.K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, Biometrika 57 (1) (1970) 97–109, arXiv:https://academic.oup.com/biomet/article-pdf/57/1/97/23940249/57-1-97.pdf, https://doi.org/10.1093/biomet/57.1.97.
[19] B. Stenqvist, A. Thuresson, A. Kurut, R. Vácha, M. Lund, Faunus – a flexible framework for Monte Carlo simulation, Mol. Simul. 39 (14–15) (2013) 1233–1239, https://doi.org/10.1080/08927022.2013.828207.
[20] A. Kurut, B.A. Persson, T. Åkesson, J. Forsman, M. Lund, Anisotropic interactions in protein mixtures: self assembly and phase behavior in aqueous solution, J. Phys. Chem. Lett. 3 (6) (2012) 731–734, https://doi.org/10.1021/jz201680m.
[21] C. Pasquier, S.R. Midtgaard, M. Polimeni, C.I. Jørgensen, L. Arleth, T.H. Callisen, M. Lund, Anisotropic protein-protein interactions in dilute and concentrated solutions, J. Colloid Interface Sci. 629 (2023) 794–804, https://doi.org/10.1016/j.jcis.2022.08.054.
[22] L.A. Delboni, F.L. Barroso da Silva, On the complexation of whey proteins, Food Hydrocoll. 55 (2016) 89–99, https://doi.org/10.1016/j.foodhyd.2015.11.010. https://www.sciencedirect.com/science/article/pii/S0268005X15301429.
[23] S. Brownlow, J. Morais-Cabral, L. Sawyer, Bovine beta-lactoglobulin, lattice x (May 1997). doi:https://doi.org/10.2210/pdb1beb/pdb.
[24] S. Moore, B. Anderson, C. Groom, M. Haridas, E. Baker, Structure of differic bovine lactoferrin at 2.8 angstroms resolution (Dec. 1997). doi:https://doi.org/10.2210/pdb1blf/pdb.
[25] A.Z.P.J. Karl Johnson, K.E. Gubbins, Reactive canonical monte carlo, Mol. Phys. 81 (3) (1994) 717–733, https://doi.org/10.1080/00268979400100481.
[26] S. Ulrich, A. Laguecir, S. Stoll, Titration of hydrophobic polyelectrolytes using Monte Carlo simulations, J. Chem. Phys. 122 (9) (2005) 094911, https://doi.org/10.1063/1.1856923.
[27] R. Bürgi, P.A. Kollman, W.F. van Gunsteren, Simulating proteins at constant ph: an approach combining molecular dynamics and Monte Carlo simulation, Proteins: Struct. Funct. Bioinf. 47 (4) (2002) 469–480, https://doi.org/10.1002/prot.10046.
[28] R. Penfold, J. Warwicker, B. Jönsson, Electrostatic models for calcium binding proteins, J. Phys. Chem. B 102 (43) (1998) 8599–8610, https://doi.org/10.1021/jp973420s.
[29] A. Warshel, S.T. Russell, A.K. Churg, Macroscopic models for studies of electrostatic interactions in proteins: limitations and applicability, Proc. Natl. Acad. Sci. 81 (15) (1984) 4785–4789, https://doi.org/10.1073/pnas.81.15.4785.
[30] Mikael Lund, Bo Jönsson, Clifford Woodward, Implications of a high dielectric constant in proteins, J. Chem. Phys. 126 (22) (2007), https://doi.org/10.1063/1.2741543.
[31] B.A. Persson, M. Lund, Association and electrostatic steering of α-lactalbumin–lysozyme heterodimers, Phys. Chem. Chem. Phys. 11 (2009) 8879–8885, https://doi.org/10.1039/B909179C.
[32] G. Tesei, Y.-W. Hsiao, A. Dabkowska, G. Grönberg, M. Yanez Arteta, D. Ulkoski, D. J. Bray, M. Trulsson, J. Ulander, M. Lund, L. Lindfors, Lipid shape and packing are key for optimal design of ph-sensitive mrna lipid nanoparticles, Proc. Natl. Acad. Sci. 121 (2) (Jan. 2024), https://doi.org/10.1073/pnas.2311700120.
[33] F.L.B. da Silva, B. Jönsson, Polyelectrolyte–protein complexation driven by charge regulation, Soft Matter 5 (2009) 2862–2868, https://doi.org/10.1039/B902039J.
[34] I. Sogami, N. Ise, On the electrostatic interaction in macroionic solutions, J. Chem. Phys. 81 (12) (1984) 6320–6332, arXiv:https://pubs.aip.org/aip/jcp/article-pdf/81/12/6320/18952657/6320_1_online.pdf, https://doi.org/10.1063/1.447541.
[35] J. Wu, D. Bratko, J.M. Prausnitz, Interaction between like-charged colloidal spheres in electrolyte solutions, Proc. Natl. Acad. Sci. 95 (26) (1998) 15169–15172, https://doi.org/10.1073/pnas.95.26.15169.
[36] Y. Li, M. Mascagni, A. Gorin, A decentralized parallel implementation for parallel tempering algorithm, Parallel Comput. 35 (5) (2009) 269–283, https://doi.org/10.1016/j.parco.2008.12.009. https://www.sciencedirect.com/science/article/pii/S0167819108001397.
[37] I. Coluzza, D. Frenkel, Virtual-move parallel tempering, ChemPhysChem 6 (9) (2005) 1779–1783, https://doi.org/10.1002/cphc.200400629.
[38] H. Wennerström, Statistical Mechanical Description of Surfactant Self-Assmbly. In Organized Solutions: Surfactants in Science and Technology, Marcel Dekker, 1992.
[39] A.I. Abrikosov, B. Stenqvist, M. Lund, Steering patchy particles using multivalent electrolytes, Soft Matter 13 (26) (2017) 4591–4597, https://doi.org/10.1039/c7sm00470b.
[40] A. Lošdorfer Božič, R. Podgornik, Anomalous multipole expansion: charge regulation of patchy inhomogeneously charged spherical particles, J. Chem. Phys. 149 (16) (2018) 163307, https://doi.org/10.1063/1.5037044.
[41] A.L. Bozic, R. Podgornik, Ph dependence of charge multipole moments in proteins, Biophys. J. 113 (7) (2017) 1454–1465, https://doi.org/10.1016/j.bpj.2017.08.017.
[42] D. Bratko, A. Striolo, J. Z. Wu, H. W. Blanch, J. M. Prausnitz, Orientation-averaged pair potentials between dipolar proteins or colloids, J. Phys. Chem. B 106 (10) (2002) 2714–2720. doi:https://doi.org/10.1021/jp013685d.
[43] M. Lund, B. Jönsson, On the charge regulation of proteins, Biochemistry 44 (15) (2005) 5722–5727, https://doi.org/10.1021/bi047630o.
[44] C. Yigit, J. Heyda, J. Dzubiella, Charged patchy particle models in explicit salt: ion distributions, electrostatic potentials, and effective interactions, J. Chem. Phys. 143 (6) (2015) 064904, https://doi.org/10.1063/1.4928077.
[45] C. Yigit, J. Heyda, M. Ballauff, J. Dzubiella, Like-charged protein-polyelectrolyte complexation driven by charge patches, J. Chem. Phys. 143 (6) (2015) 064905, https://doi.org/10.1063/1.4928078.
[46] C. Yigit, M. Kanduč, M. Ballauff, J. Dzubiella, Interaction of charged patchy protein models with like-charged polyelectrolyte brushes, Langmuir 33 (1) (2016) 417–427, https://doi.org/10.1021/acs.langmuir.6b03797.
[47] K. Olechnovic, C. Venclovas, Voronota: a fast and reliable tool for computing the vertices of the voronoi diagram of atomic balls, J. Comput. Chem. 35 (8) (2014) 672–681, https://doi.org/10.1002/jcc.23538.
[48] A. Bowyer, Computing Dirichlet tessellations*, Comput. J. 24 (2) (1981) 162–166, arXiv:https://academic.oup.com/comjnl/article-pdf/24/2/162/967239/240162.pdf, https://doi.org/10.1093/comjnl/24.2.162.
[49] A. Poupon, Voronoi and voronoi-related tessellations in studies of protein structure and interaction, Curr. Opin. Struct. Biol. 14 (2) (2004) 233–241, https://doi.org/10.1016/j.sbi.2004.03.010. https://www.sciencedirect.com/science/article/pii/S0959440X04000442.
[50] W. Humphrey, A. Dalke, K. Schulten, VMD – visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38.
[51] X. Huang, R. Pearce, Y. Zhang, Faspr: an open-source tool for fast and accurate protein side-chain packing, Bioinformatics 36 (12) (2020) 3758–3765, https://doi.org/10.1093/bioinformatics/btaa234.
[52] A.Z. Panagiotopoulos, Direct determination of phase coexistence properties of fluids by Monte Carlo simulation in a new ensemble, Mol. Phys. 100 (1) (2002) 237–246, https://doi.org/10.1080/00268970110097866.