UNIVERSITY OF TURKU

# Protein condensates: connecting experimental phenomena with computational methods

Materials Engineering

Bachelor's thesis

Kaisa Alasaari

15.05.2026

Turku

Bachelor's thesis

**Subject**: How do different modelling scales connect to experimentally observed phenomena in protein self-assembly and condensates

**Author(s)**: Kaisa Alasaari

**Title**: Protein condensates: connecting experimental phenomena with computational methods

**Supervisor(s)**: Narayan Gupta

**Number of pages**: 24 pages

**Date**: 15.05.2026

Protein condensates have been proven to be the main cause for multiple different diseases. Although a lot of research have been done, the key reasons for formation have not been found, making it impossible to develop treatments and drugs to prevent these diseases. While traditional methods for imaging can be useful, especially when used with other, new methods, simulations have improved the speed and accuracy of our knowledge of condensates. In this thesis we will discuss the main pathways to the solidified condensates: from droplet formation to gelation. In addition, we will explain what the pathological pathways to toxic condensate formation are and why the cellular regulations can be powerless against them. Understanding these natural processes and reasons why some of them escape the cellular regulations could in theory help up fight against neurodegenerative diseases. As mentioned, computational methods are often required to achieve the necessary time and size scales of interactions that affect the condensation formation. Multiple methods have been created for different scales, and they all contribute to unravelling the secrets behind protein condensation.

Proteiinikondensaattien on todettu olevan syy useiden sairausten esiintymiselle. Tutkimustyön laajuudesta huolimatta on edelleen epäselvää, minkä vuoksi kondensaatteja muodostuu. Tämän epäselvyyden vuoksi kyseisiä sairauksia ei voida parantaa, eikä niihin ole mahdollista kehittää sairauksille altistumista estävää lääkitystä. Moni perinteinen menetelmä on äärimmäisen hyödyllinen kondensaattien tutkimuksissa ja suurin hyöty saadaan yhdistelemällä useampaa menetelmää. Uudet simulaatioita hyödyntävät mallinnusmentelmät ovat kuitenkin kyenneet vielä tarkempaan ja nopeampaan tutkimustyöhön. Tässä tutkielmassa käymme läpi kondensaattien muodostumisen pääväylät pisaroitumisesta geeliytymiseen. Lisäksi käymme läpi lyhyesti patologisten kondensaattien muodostumisen ja syyt näiden muodostumien kykyyn välttää solun omat säätelykeinot. Kuten mainittu, mallinnusmenetelmien avulla on mahdollista parantaa sekä aika- että kokoskaalaa huomattavasti perinteisiin menetelmiin verrattuna. Nämä menetelmät ovat räätälöity tietylle kokoluokalle, mutta kaikki ovat yhtä tärkeitä kondensaattien muodostumisen syiden selvittämistä varten.

**Keywords**: biological molecule, coarse-grained, computation, condensate, dynamics, experimental, protein

# Table of contents

# 1 Introduction

Neurodegenerative diseases such as Alzheimer's, Parkinson's and Huntington's share a common cause: abnormal protein condensates. Something goes wrong during the protein folding, or proteins begin to clump up in patterns that they are not supposed to [1]. These events lead to unnatural accumulation of structures, that are not supposed to form. Multiple different pathological pathways have been identified [1] and scientifical evidence points at the direction of protein condensates. While we have found a cause for these diseases, which are all currently uncurable, we have not found a reason for the appearance of the condensates.

Protein self-assembly as a phenomenon is not itself unnatural: the cell has multiple different organs that are formed by the same process [2], even protein condensates can have a biological function. In fact, this tendency seems to be naturally occurring to proteins themselves, not just inside a cell [3]. Multiple studies have been conducted, but the dynamical nature of the condensates have made it difficult to uncover the reasons behind the formation of these structures. Microscopy has been an invaluable tool to identifying which proteins tend to undergo condensation, but the resolution of the method is lacking both time and size scaling.

To overcome these issues, computational methods have been introduced. In simulations, it is much easier to define the areas of interest and focus on smaller details. In fact, we have multiple discoveries of protein behaviour and what attributes affect condensation the most  by using these methods [1,4]. However, even though methods and processes have improved drastically over the years, we have still yet to answer the burning question why these unnatural condensates are forming.

To fully comprehend the phenomena explained in this thesis, we must first understand the underlying process that causes these condensates to form.

Proteins have multiple ways of forming large structures, one of which is liquid-liquid phase separation (LLPS) [6], where two distinctively different phases appear. A simplification of LLPS is depicted in figure 1. A denser area of proteins begins to form within the solution (dark green), making the surrounding area less dense (light green). If the concentration of the proteins is increased, the dense phase becomes more prominent in conditions where phase separation can occur.

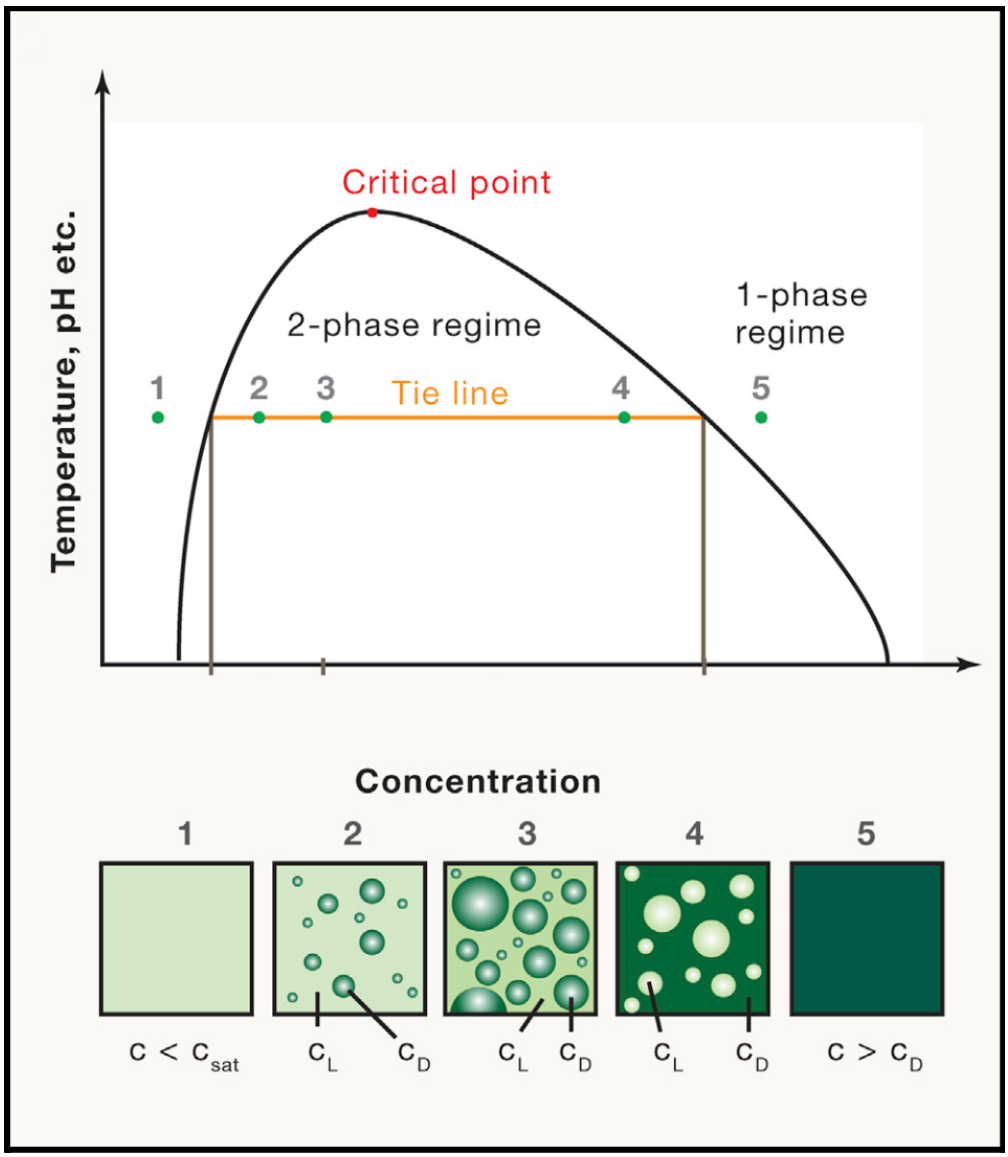


Figure 1 Phase diagram of a molecule. Y-axis represents any condition that can affect phase separation. At 1, concentration is not high enough for phase separation to happen. At 2-4, two distinct phases can be observed: light phase ($c_L$) and dense phase ($c_D$). At 5, the whole phase is considered dense phase. Phase separation cannot occur beyond the critical point. Modified from Alberti et al. [5]. Copyright 2019 Alberti, published by Elsevier Inc

Phase separation is simplistically caused by proteins attracting each other and bundling up into small droplets, while thermodynamically acting as a liquid [4]. Unlike a mixture of oil and water where oil also forms small droplets, condensates formed through LLPS can and will exchange material with the surroundings. Most commonly this material is nucleic acids and possibly other proteins with similar characteristics [6]. Based on experiments, multivalency and proteins containing intrinsically disordered regions (IDR) have been declared most important factors to initiating LLPS [2]. Multivalent proteins have multiple binding sites, making them act like glue [6]. IDRs are string-like regions, that don't have a stable form [2]. They will highly increase the flexibility of a protein and can adapt to almost any conformation [2]. This dynamic form also changes what kind of interactions the protein has with other molecules.

What makes this pathway so interesting is that the densely condensed areas of proteins can form in biological systems and synthetically [3]. They also have multiple different states and each of these states behave differently and react quite drastically to even the smallest changes in cellular conditions such as pH, temperature or kinetic energy [3,4]. Characteristics of these states are further explained in section 2.

This thesis aims to give insight into what kinds of phenomena have been observed in systems that form condensates, how these phenomena form and what kinds of tools we currently have to both observe and simulate these phenomena. We will also briefly discuss current issues regarding computational methods. Going forward, *bonds* will refer to specifically non-covalent bonds and *solution* will refer to any liquid solution containing proteins.

## 2 Experimental phenomena in protein self-assembly

In this section we will go through different phases of condensate formation. Like the name suggests, experimentally observed phenomena can be anything that is observed during live solution experiments. By altering the conditions of the solution, such as temperature or pH, the condensate might behave differently. Proteins with distinctly different conformations and compositions will most likely also behave differently under same conditions. Therefore, it is quite difficult to apply a general rule that covers all proteins, but this section will focus on the most common types of proteins that exhibit the tendency to form condensates and their phenomena. In this thesis, *droplet* will refer to the physical state of the condensate, rather than a specific size of the condensate.

Figure 2 presents these phenomena in a visual format. The circles inside the condensate do not mean droplets themselves, instead they are a simplification of a protein.

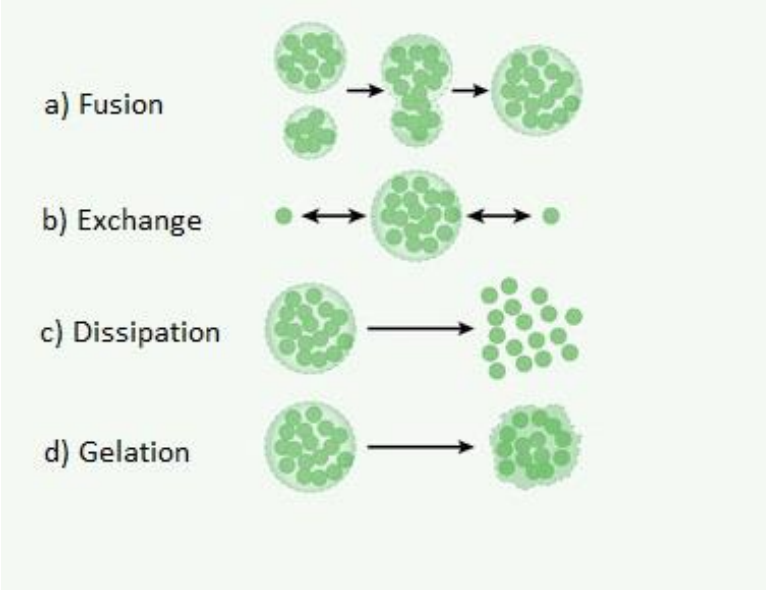


Figure 2 Experimentally observed phenomena of condensates, a) depicts the fusion of two droplets, b) depicts dynamic exchange of material with the solution, c) depicts the dissipation of the droplet and d) depicts the aging and gelation of a droplet. Modified from Jiang et al. [7]. Copyright 2022 Jiang, published by Wiley

## 2.1 Condensate formation

### 2.1.1 Droplets

In the early stages of liquid-liquid phase separation, condensate formation begins with initiating macromolecules called scaffolds [12]. While protein condensates can contain multiple different macromolecules, studies have shown that only a select few will initiate the condensation process [13]. Scaffolds have a notable feature of multivalent domains [10] and will form the initial droplet and can recruit other molecules, called clients, of lower valency into the condensate [13]. Clients do not typically promote condensation but will form bonds with scaffolds and other clients [13]. Those with higher valency will be more strongly attracted to scaffolds [13]. Just like in condensation forming, scaffolds and their ability to attract clients is extremely sensitive to change; small shift in scaffold concentration or the stoichiometry of the molecule can cause extensive changes in client recruitment [13]. Electrochemical interactions are one of the most important interactions between molecules in the initial droplet formation [2]. These forces form between two oppositely charged regions, and compared to other interactions such as pi-pi interactions, the effective range of the interaction is significantly higher [2], thus promoting initial droplet formation in a mixed state.

Within condensate-associated proteins only some regions are actively promoting the droplet formation. Therefore, a single polymer structure can be divided into two separate regions. First, a sticker region, where the polymer can form bonds with itself (intrachain) or other polymers (interchain) and second, a spacer region that has a lower tendency to form bonds [11]. A sticker can be a structured domain, or an amino acid that has a particularly strong affinity to form a bond, typically by polarity or aromatic interactions [11]. A spacer on the other hand is a region where the structure of the protein does not form bonds as readily as the stickers [11]. Spacers are often also noticeably more flexible regions, and they will often be composed of IDRs [11].

Droplets are highly dynamic condensates [8]. They form and dissipate constantly, and they possess extreme adaptivity to environmental changes [8]. This adaptivity is partially due to their ability to exchange materials with the surrounding environment as well as other droplets [9]. The exchanging of material with the surrounding solution is what

makes them reversible: the formed structure can be freely dismantled. Droplets will form into a sphere. Just like any liquid, condensate droplets operate under the laws of fluid mechanics and will form a spherical shape due to surface tension. Even though the condensate is enclosed within a solution, the interactions with the other members of the condensate will be stronger than the interactions with the solution, thus creating a sphere.

### 2.1.2 Arrest and fusion

A molecule has a limited number of binding sites. If these sites are available to other surrounding molecules, they might form a bond. However, if a cluster of these molecules is oriented in a way that all the binding sites are turned in towards the centre of the cluster, it cannot form bonds with the surrounding environment. This lack of bond forming is called arrest [9]. A droplet does not need to occupy all its binding sites to undergo arrest, rather a droplet is said to undergo arrest when it can no longer merge with another droplet [9].

Like other liquids, droplets can fuse with each other if they are near one another. Due to the condensate forming via weak interactions, the internal structure can change and form a new structure with the absorbed molecules. While a high protein density within the droplet will correlate sightly with the ability to fuse, it is not definitive. Clusters with low density can undergo arrest while a high-density cluster might fuse easily [9]. Another, quite highly correlating feature of droplets is the number of reorganised bonds within the droplet. To break a sticker-sticker bond, a certain amount of energy is needed: the same amount of energy as is the strength of the sticker interactions ($E_s$) [9]. For a droplet with a large $E_s$, the number of reorganised bonds will be much lower than for a low Es droplet [9]. Interestingly, it is rather impossible to predicts a droplet's tendency to undergo fusion by looking at its static properties alone such as density or number of free stickers. Instead, the lifetime of a bond – and in turn high $E_s$ – is a key factor for cluster fusion [10]. These phenomena are true for a mixed solution only. Simulations have shown that droplet's density causes arrest in homopolymers that are composed of spacers only, and the dynamic behaviour of the bonds inside the cluster cause arrest in sticker-spacer heteropolymers [9].

### 2.1.3 Aging and gelation

Aging in condensates refers to the loss of dynamics: the condensate will gradually exchange less material with the surrounding solution, and it will slowly start to solidify into a gel over a period of time [2]. This loss of dynamics also means that the condensate will no longer dissipate as readily compared to droplets. Depending on the condensate composition, gelation can be a reversible stage [2]. Gelation will most commonly happen through phase separation, where the entire droplet will be affected [2,11]. The proteins will form a system-wide network [11] that increases the stiffness and viscosity of the condensate. Although it is possible for gelation to occur without a droplet formation first, it is rather rare and most of the biological functions favour gelation through LLPS [2]. From a biological perspective this makes sense, since the cell has more control over droplets' formation and dissipation. As gelation affects the whole system, gelation without LLPS would affect the whole cell.

Just like with droplets, the timescales for bonding and bond breaking will also affect a gel's flow and deformation [10]. Other properties such as stiffness and porosity are affected by the number of bonds and therefore proteins recruited [10].

## 2.2 Pathological pathways

Under normal conditions, LLPS is ubiquitous and causes biological condensates to form and dissipate. This delicate balance is usually under strict cellular regulation [12]. However, environmental changes or mutations can cause abnormalities in the condensate formation process: for example, phase transition from liquid to solid state will cause changes to the condensate's mechanical and chemical properties [8]. If a condensate undergoes an abnormal phase transition, it might gain toxic properties or lose its original function [8]. If this phase transition is irreversible – as is usually the case in aggregates- the cell cannot regulate the formation or dissipation of the condensates[12] [8]. Protein condensates most commonly associated with diseases are proteins that also have a functional role in the cell [12]. Most diseases caused by condensate formation affect the aging population, which indicates that the irreversible

changes within the cells need time to accumulate before causing the problems associated with them.

## 2.3 Imaging methods for experimentally observed phenomena

Due to the size scale of the experiments, it almost always necessary to use an imaging method to observe the changes in the solution. In some cases, it is possible to see the condensation with a naked eye, although this type of data is rarely useful. The most common type of imaging is light microscopy [12]. It is highly adaptable and often not as expensive as more advanced options. Microscopy is an excellent choice for determining whether condensation happens at all, and how large clusters are forming after a specific time [12,13]. The number of formed clusters can also be calculated quite reliably. Although microscopy is rather slow, it is well suited for qualitative experiments of biomolecular condensates. The most interesting part of condensate forming cannot be captured by pure microscopy: dynamics. Another drawback of pure microscopy is phototoxity in live cells. UV-light can damage cells by ionisation, and visible light can react with compounds in the medium or the cell itself. However, multiple different methods have been invented [12] to solve these issues.

### 2.3.1 Fluorescence recovery after photobleaching (FRAP)

FRAP is an imaging method that works by exploiting the fluorescent molecules' tendency to switch off (*bleached*) after extremely intense light exposure. The molecules of interest are tagged with a fluorescents and precisely targeted with corresponding laser light to bleach only tagged molecules [12]. After bleaching, pictures are captured over time to follow the recovery of the fluorescent molecules [12]. The slow increase in emitted light means that the position of the molecule can be accurately tracked within the condensate, allowing for the determination of the diffusion kinetics. FRAP can be used in both *in vitro* and live cells. Phototoxity can be prevented by the accurate laser light, thus enabling the imaging of the same sample multiple times. FRAP is typically used to follow only one type of proteins, so condensate properties should be taken into consideration before drawing conclusions of the material properties. Different proteins

in a condensate can have different diffusion rates, which will affect the calculated viscosity of the condensate. However, just like with microscopy, getting an accurate 3D image of the sample is extremely difficult.

### 2.3.2 Digital holographic microscopy (DHM)

DHM is an advanced technique that utilises wave interference to construct accurate holograms of samples. A laser beam passing through a transparent (or semi-transparent) sample is recorded and combined with a refence wave coming from the same source. This combined wave is then reconstructed into a hologram by appropriate software. A 3D shape of the object is created by moving the object, and its internal structure can be captured for transparent samples. DHM is an excellent tool for live cells, since they are often transparent and need staining to be observed with more classical methods. DHM can be used on an extremely small scale and is therefore very useful to condensate studies. DHM is also a real-time method, which is very important when observing changes within the condensate [14]. Due to its ability to pass through cellular material or growing mediums, it is also very capable of accurately depicting multiphase liquid condensates.

Figure 3 is an example of a DHM image. It contains multiple HeLa cells, which are cancerous cells from cervix [15]. DHM captures the borders of each cell's nucleus (lighter spots, blue arrow) and inner condensate formations within the nucleus (white spots, red arrow).

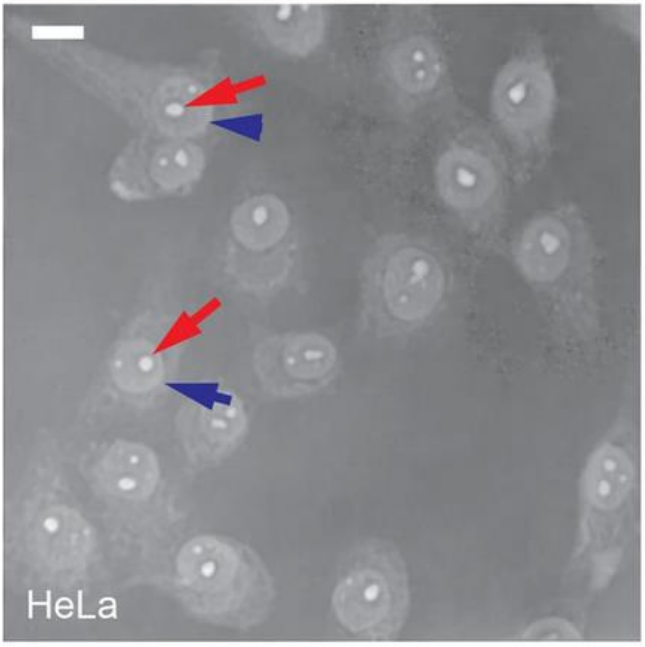


Figure 3 HeLa cells captured by DHM. Blue arrow indicates the edges of the nucleus, while red arrow indicates the internal structures within the nucleus. Modified from Zorbas et al. [15]. Copyright 2024 Zorbas, published by Nature Publishing Group UK

# 3 Multiscale computational modelling

Computational methods provide a way to analyse the molecules on a level of detail that would be impossible to achieve with experimental methods [12], as well as allow the running of multiple simulations with slight differences at the same time. Simulations are also an easy way to gather data directly and in large quantities. Computational methods can also be used to predict certain outcomes in specific conditions [12], which would be an invaluable tool in healthcare. In figure 4 is depicted the typical range of computational methods.

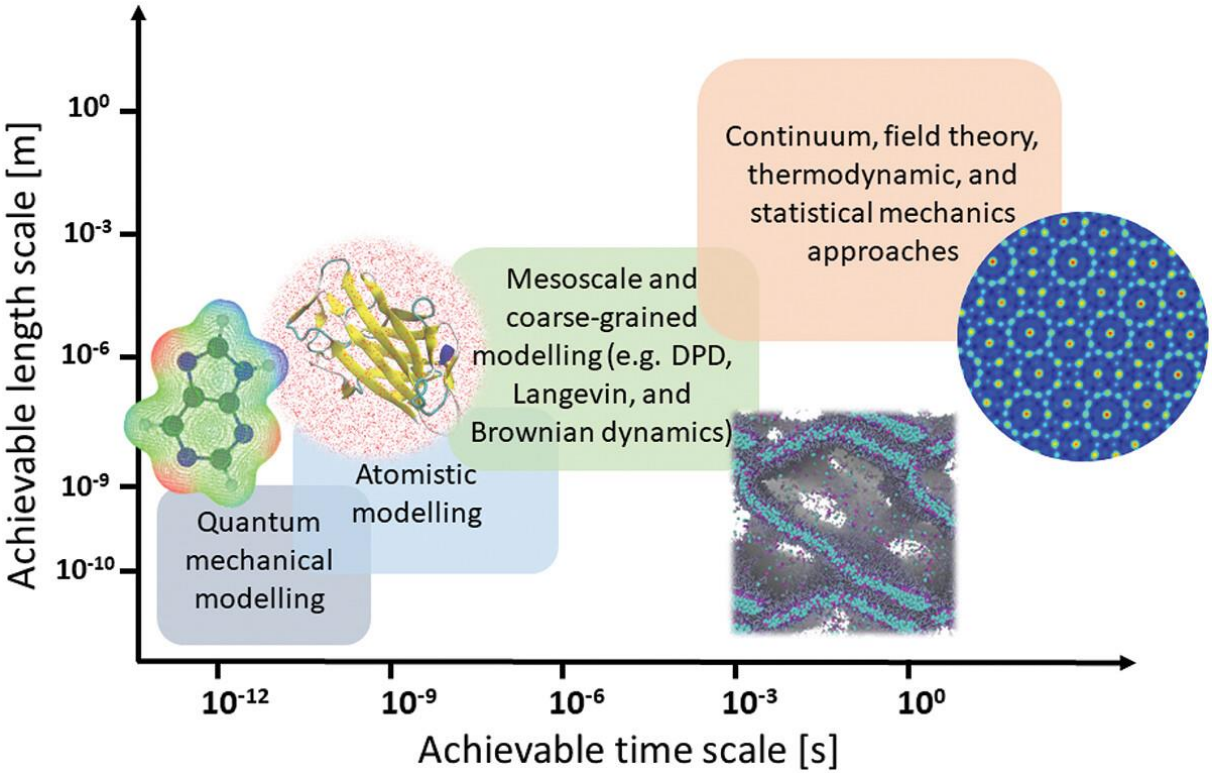


Figure 4 Different computational methods visualised on a time and size scale. Modified from Scacchi et al. [4]. Copyright 2024 Scacchi, published by Taylor & Francis

The molecules that form these condensates are extremely complex, and typically have regions that interact at numerous different timescales [12]. Common timescales range anywhere from nanoseconds to minutes, making system-wide simulations of atomic levels of detail impossible to execute. Instead, each computational method has its own level of detail and corresponding timescale. Just like in any research, it is important to

recognise which interactions are crucial to the problem at hand, and which method can provide the necessary data efficiently and accurately.

Physics based computational methods use force fields and water models [12] to improve the accuracy of the models [12]. Force fields will reproduce the molecular properties derived from either experimentally gained data or quantum mechanical calculations [12]. Each method will typically have its own specific models to improve efficiency and specificity of the method.

## 3.1 Atomistic models

Atomistic models can be divided into two distinct methods: quantum mechanics and molecular mechanics [8]. While the molecular mechanics is much more utilised due to lower costs and higher number of simulated atoms [1], the quantum mechanics model is invaluable tool that can be utilised to determine parameters for molecular mechanics. The main limiting factors for modelling biomolecular condensates are the limited size of the structure as well as the timescale. While singular bond formation can be captured, large scale structures form too slowly for atomistic models [4]. Overall, atomistic models are great at predicting and simulating conformational changes within the molecules.

### 3.1.1 Quantum mechanics

The quantum mechanics model is best suited to the study of a system consisting of hundreds to thousands of atoms. It is incredibly accurate at describing spectroscopic details [1] and bond rearrangement energetics [16]. Quantum mechanics model can be optimised to different scopes of detail by altering the extent of mathematical estimation (extrapolation) [1], level of theory applied and the amount of computational resources. At the highest level of detail quantum mechanics can offer detailed overview of individual amino acid potential energy surface, and at the lowest level predict a protein’s folding pattern [1].

However, quantum mechanics suffers from extremely limited timescales, as well as high costs. Therefore it is traditionally used to gather data for other models [1], and with molecular mechanics. When used with molecular mechanics, quantum mechanics model can be used to specifically simulate interactions that cannot be done on molecular mechanics' scale [16], e.g. charge transfers between atoms.

### 3.1.2 Molecular dynamics (MD)

Molecular dynamics is very similar to quantum mechanics but has slightly higher scale: it has a size limitation up to 30 $nm^3$ and timescale of μs [4]. Benchmark simulations of billion atoms and 100 ns in time exist, although this has been possible only with an extremely optimised setup [4]. Molecular dynamics is very popular, due to its capability to simulate dynamic and structural data at atomic level [1] and more flexible range than quantum mechanics. Another key feature of molecular mechanics is the simulation of oligomeric interactions in liquids [1].

While more practical than the quantum mechanics model, molecular mechanics does have downsides. Most notably, the lack of formation of highly ordered structures and phase separation due to time and length limitations [1,4]. Other key issue is the verification of the model's accuracy [4]. Simulated interactions are near impossible to confirm vie experiments, due to their extremely small scale.

## 3.2 Coarse-grained models (CG)

Coarse-grained models have a larger scope than atomistic models, and in turn have less details [12]. This method can simulate structures up to mm in length and tens of ms in time. In coarse-grained model multiple atoms are grouped into a single bead. The number of atoms chosen to represent a residue can be changed, depending on the protein of interest and what type of behaviour is studied. This allows for a large effective range for both the time and size scaling for the method. Coarse-grained methods are typically labelled as physics based, as they are often based on equations of motion.

Although the atoms are reduced into a bead, the structure can be simulated to interact like the original molecule. In a way, the molecule loses its shape but not its nature. It will still want to bind with certain other molecules. This interaction strength is an average of the values gained from other methods, such as atomistic models [4]. These interactions are added as binding sites to the surface of the bead, which allows for accurate multivalency behaviour to be captured.

### 3.2.1 Coarse-gained methods for molecular dynamics

Molecular level of CG is the smallest. It can achieve sizes up to 10 nm and timescale of μs and beyond [4]. CG MD is based on integrations of Newton's motion equations, making it physics based. These equations allow for calculations of the kinetics and thermodynamics of a system [4].

Two main, broad-scope force field approaches are top-down and bottom-up based MD field models. Top-down is thermodynamics based and suited for accurately reproducing experimentally observed phenomena such as thermodynamic properties [17]. It will also produce very accurate interaction potentials that are easily transferable to the target system. Bottom-up is based on atomistic detail gained from MD simulations [17]. It is very good at retaining the target system's chemical characteristics. These two field models are often used in a hybrid system to fine-tune the coarse-grained model's parameters [17].

When it comes to specifically biomolecules, Martini is the most common [4]. It is easily transferable to multiple different molecular systems and has simplistic interaction parameters. These bead interaction include non-bonded, derived from both bottom-up and top-down approaches, and bonded, which are optimised to represent data gathered from MD methods [4]. Multiple specialised Martini models have been created to better suit the specific systems, but the lack of absolute repulsive interactions in the model's formation makes it incapable of modelling soluble proteins and equilibrium between two systems [4].

Issues for coarse-grained methods in MD include the dependence on surrounding chemical environment, and challenging simulations of highly localised phenomena

caused by loss of chemical detail [4]. Examples of localised phenomena are ion effects, changes in binding sites and molecular recognition.

### 3.2.2 Dissipative particle dynamics (DPD)

Dissipative particle dynamics extends the CG MD model's range significantly both in size and time. This makes it an excellent model for studying the structure assembly and changes in morphology over time [4]. DPD simulates the fluid dynamics and liquid responses of a system by examining the time evolution of the system's particles [4]. It is also based on Newton's motion equations. DPD has a soft-core potential, allowing for much longer timescales as well as improved relaxation dynamics. Standard DPD considers all interacting beads to be of the same size and mass [4].

Although DPD is very powerful, it has some drawbacks when working with biomolecules. Standard DPD has no dissipative force, meaning no shear force between two particles. Liquid-vapour interfaces cannot be modelled either, though both issues have been remedied by alternative models, such as many-body DPD. With the addition of polarizability, DPD can accurately replicate native protein structure and folding pathways [4].

### 3.2.3 Langevin and Brownian dynamics

Unlike the previous CG methods, which are based on Newton's laws of motion, Langevin and Brownian dynamics (BD) are based on numerical techniques capable of simulating equilibrium and non-equilibrium phenomena [4]. Both methods simulate the system's dynamics in the presence of a solvent, however, Langevin is optimised for viscous solvent and its effects on a system [4]. These two methods achieve even higher levels of coarse graining and in turn longer lengths and timescales.

Langevin dynamics has a remarkable capability to map both dynamics and assembly of a system [4]. It can also be used to study protein sequence's effect on phase separation. BD on the other hand can be used to map proteins' dependency on charge regions and in turn differentiate different interactions such as non-specific binding and local charges [4].

#### 3.2.4 Monte Carlo (MC)

Monte Carlo generally refers to any simulation where the system's response is modelled after random sampling of all the states that the system can take. This outcome can be weighted towards specific distributions, such as Boltzmann. While Monte Carlo is based on CG methods and the sampled states are also based on physics, it can be labelled as data driven due to the random sampling method. MC can be used to study fibril formation and dynamics. Fibrils are the determined main cause for multiple neurodegenerative diseases and consist of repeated patterns of structured proteins [17]. It can also simulate morphological differences in condensates caused by the proteins' conformational differences. MC is a powerful tool to predict biomolecular systems' structures.

### 3.3 Patchy particles

As discussed previously in section 2, the polymer structure can be divided into two different parts: stickers and spacers. This distinction can be utilised with coarse graining of the molecule, producing a minimal model of the condensate that can model the dynamics accurately. Unlike the DPD methods, patchy particles use a hard core with attractive anisotropic sites called patches [18]. These directional interactions have been found to be critical to the phase separation of a condensate [18]. Accurate reconstruction of phase separation behaviour is also crucial to simulating changes in the phase separation when studying mutated proteins [18]. Patchy particle methods have also helped in the development of protein manipulation strategies, such as high quality crystal production [18].

The directional attraction of the particles is the defining feature of the patchy particle method, and multiple different ways for modelling exist. Just like the other models, multiple parameters, such as patch orientation, range and angular width [18], can be altered to better suit the problem at hand. Multivalent particles can be modelled by allowing a single patch to form multiple bonds [18]. The directional attraction makes the patched more selective and lowers the absolute amount of bonds formed. This allows

accurate modelling for structures that do have binding sites but suffer from steric hindrance.

Patchy particles can be modelled by using MC sampling [18]. This allows for the simulation of bond probabilities by simulating all the possible moves and rotations that a particle can take in the specified space [18]. Other simulations include critical point estimation (where phase separation occurs) and phase coexistences for vapour-liquid and fluid-crystal phases [18].

### 3.4 Mean field models

Mean fields are a collection of models that provide solvable models for complex problems, such as phase separation [19]. The size range is between micro and macro scale, allowing the study of material flow and large-scale reorganisation [4]. Molecular level structure and molecules' particle-like characteristics are however lost.

One of the most used models utilises the Flory-Huggins solution theory, which excels at providing insight about density transitions. According to the Flory-Huggins theory, the competition between interaction free energy and entropic free energy is the driving force of phase separation [19]. A simplistic equation for free energy density is presented in equation 1,

$$f(\phi) = \frac{\phi}{N} * ln\phi + (1-\phi) * \ln(1-\phi) + \chi\phi(1-\phi) \quad (1)$$

where ϕ is polymer volume fraction, N is the polymer chain length and χ is the parameter for solvent-polymer interaction [19]. By utilising this simple equation, it is possible to predict the phase separation [19]. This equation can be modified to include the addition of e.g. surface tension, which will cause disfavouring of density fluctuations [19].

This model will consider the conditions of the phase separation to happen at thermodynamical equilibrium, that is, the energy is conserved during reactions. However, that is not the case with cellular conditions as the cell includes multiple energy consuming processes. These processes can be included in the models by the

addition of dynamic processes between solutes [19]. In the case of proteins consuming energy directly, this activity become a characteristic of the solute and needs to be included by, for example, adding a nonintegrable parameter to the free energy function [19]. Although mean field methods don't have the details for chemical specificity of the system, they are extremely good at modelling structural changes that happen on larger timescales that cannot be reached by other methods [19].

# 4 Connecting experimental phenomena and computational methods

As discussed earlier in section 3, each computational method has its own advantages and disadvantages. Tables 1 to 4 describe examples for each computational method, although some of the simulations have utilised additional methods such as FRAP to confirm and further refine gained knowledge. These are only examples; each method has much more uses than described here. The use of two computational methods for a singular problem is extremely common, due to the limitations of a singular system. Apart from nucleation events, all observed phenomena will happen on a scale larger than what one method can achieve.

Atomistic models are best for localised problems and determining material properties that are related to chemical properties, for example melting point can be calculated from bond energy. This value would be near impossible to gain from any other method, since they lack the chemical nature of the condensate. Table 1 describes few other uses for atomistic models. Due to the limited scale, these examples focus on small scale phenomena. In theory, solvent distribution around the proteins can be crucial in any phenomena, not just nucleation and fusion.

Table 1 Experimental phenomena for atomistic models

| | *Focus of study* | *Experimental phenomena* |
|---|---|---|
| *Atomistic model* | Electrostatics [4] | Nucleation |
| | Bond angles [4] | Fusion |
| | Solvent distribution [4] | Nucleation / fusion |
| | Material characteristics [4] | - |

For coarse-grained methods, the simulated phenomena are naturally more focused on bigger structures. Although not mentioned in table 2, all levels of coarse graining can be used to study phase separation behaviour in alternating conditions. Table 2 includes examples of all levels of coarse graining previously mentioned. In general, coarse-grained methods focus on the dynamics within a system. For this reason, many of the studies are focused on fusion, arrest and aging. Crowding refers to the loss of dynamics caused by other surrounding molecules.

Table 2 Experimental phenomena for coarse-grained methods

| | *Focus of study* | *Experimental phenomena* |
|---|---|---|
| *Coarse-grained, MD* | Mechanical properties [4] | - |
| | Nucleic acid sequence effects [4] | Droplet formation |
| | Fibrillation [4] | - |
| *DPD* | Bulk self-assembly [4] | Gelation |
| | Morphological changes [4] | Arrest / gelation |
| | Aggregation [4] | Arrest |
| *Langevin* | Crowding [4] | Fusion / arrest |
| | Diffusion [4] | Fusion |
| *BD* | Cross-link density [4] | Arrest |
| *MC* | Droplet properties [4] | Droplet formation |

Patchy particles have the unique feature of adjustable, directional interactions. This feature is especially useful in situations, where having a non-directional interaction could change the amount of bonds formed. For this reason, patchy particle methods are very useful in situations where the amount of bonds formed is critical. Table 3 includes other uses for patchy particles. Density calculations are crucial to determining critical points for condensates, meaning conditions where phase separation is possible (see figure 1). Coexistence refers to the existence of two different phases, like vapour and liquid.

Table 3 Experimental phenomena for patchy particle models

| | *Focus of study* | *Experimental phenomena* |
|---|---|---|
| *Patchy particles* | Density calculations [18] | Droplet formation |
| | Coexistence [18] | - |

Mean field models have very different working principles compared to the rest of the models, and in turn have more specialised use cases, as seen in table 4. Mean field has multiple ways of including the energy consumption of the cell, so for realistic representations of the cell it is one of the best. Other uses include formation of large

structures, including smaller structures inside larger complexes. Interestingly, mean field methods offer great insight into fusion as well, seeing as they can capture reorganisation. Phase separation prediction is incredibly efficient at simulating small changes in the conditions. Mean field methods are also capable of predicting an entire phase diagram for a specific molecule [19].

Table 4 Experimental phenomena for mean field methods

| | *Focus of study* | *Experimental phenomena* |
|---|---|---|
| *Mean field* | Structure formation [19] | Gelation |
| | Reorganisation [19] | Fusion |
| | Response prediction [19] | - |

Strictly limiting a single experimental phenomenon to a single method is hard and would unnecessarily limit the creativity of scientist, but a general guideline can be created. Nucleation and its dynamics are best captured by atomistic models, due to the level of detail achievable. Droplet formation and fusion gain most information from methods that include the liquid nature and dynamics, such as DPD and Langevin / BD. Droplet arrest is highly driven by bond dynamics, so both atomistic model and BD offer great value. Aging is mostly caused by slowly fading dynamical exchange, which is best captured by any of the mid-levelled coarse-grained methods. Gelation needs a larger scale both in time and size, so data-based methods like MC of mean field can be great tools.

# 5 Current issues

Although computational methods have made great strides both in technology and variability, there are multiple issues that limit the development of further understanding of phase separation and the forces behind it. Efficiency has been improved tremendously over the years, but one of the key limitations in research is still the cost of a process compared to the usefulness of the data gained from said process. Cellular research does not increase profits directly, even if it is possible to develop drugs or medical procedures based on research. This will highly limit the available funds for biological research.

First issue comes from moving from one timescale to another. The available time scales have expanded during the recent years, and existing models can reach even larger ranges, but one of the ongoing issues is bridging the gap between two models [4]. While the timescales can be seemingly close, the difference between 10 μs and 1 ms is too great to overlap two methods without losing data. Problems arise from efficient use of resources, as it is not feasible to attempt to increase the timescale of e.g. atomistic models just to bridge the gap to CG MD. Recently, advances in machine learning have made it a promising tool for increasing efficiency, especially for models that utilise large amounts of data [12].

Another issue is the complexity of the condensation process, in terms of available calculations. Mean field models can include the energy consumption of certain processes, but it is still impossible to fully capture the dynamics within a cell [12]. Since the cell doesn’t operate within a thermodynamic equilibrium, it is not possible to combine equilibrium calculations to a model and achieve a realistic view of the cell. Cellular conditions also include multiple active proteins and complexes that will affect the condensation process, often in ways that cannot be purely calculated. While can be possible to calculate an average disruption value, the condensation process is so delicate that even slight variations from this average can cause opposite reactions in a live cell.

Finally, computational methods are, at their core, software and data. This mean the standardisation of certain processes  is necessary in order to create equally

comparable data [20]. While possible, it is not efficient to not utilise data gained from another study, but to compare the datasets, standardisation is critical. This standardisation would also help with comparisons between data gained from computational methods and experimental methods. Without inclusion of the experimentally gained data, it is possible to create models that fulfil all the given equations and theories but cannot capture realistic conditions. This requirement for confirmation of experimental studies is called validation [20], and it is mandatory in most industrial processes. Large amounts of data also require infrastructure.

# 6 Conclusions

The methods for studying protein condensates have improved drastically, and many of the underlying reasons for the formulation have been uncovered. However, condensates are still a bit of a mystery, even after extensive studies and many years of research. Although the formation process from droplet formation to gelation seems to have revealed itself, we have yet to get the answer to why this happens, and what specifically causes the fibrillation formation in neurodegenerative diseases.

Computational methods have offered us a way to calculate and predict the formation of condensates in multiple proteins based on fewer examples, and while some of these have been verified, it is still difficult to accurately prove and recreate simulations in real life. After all, it is possible for simulations to capture a system that is not possible in nature, due to difficulties in accurate cellular condition rendering. Other issues come from uniform data and accessibility of the methods. Due to the large number of different proteins, it would benefit every group greatly to have standardised methods for testing and data comparisons.

Despite these issues, researchers are constantly improving the models and finding ways to increase accuracy in issues such as non-equilibrium dynamics within a cell. Both field and water models get constant updates to keep up with the increasing knowledge, and new methods have begun to implement machine learning and AI tools to increase efficiency. Due to the large amount of data needed to process, these methods will most likely turn out to be extremely useful.

## References


1. Bhattacharya, S. & Thompson, D. Recent Advances in Mapping Protein Self-Assembly and Aggregation for Common Proteinopathies. *Acta Physica Polonica A* **145**, S37–S37 (2024).
2. Zhu, Q. *et al.* Biomolecular Condensates as Emerging Biomaterials: Functional Mechanisms and Advances in Computational and Experimental Approaches. *Advanced Materials* **37**, e10115 (2025).
3. Yuan, C., Li, Q., Xing, R., Li, J. & Yan, X. Peptide self-assembly through liquid-liquid phase separation. *Chem* **9**, 2425–2445 (2023).
4. Scacchi, A., Vuorte, M. & Sammalkorpi, M. Multiscale modelling of biopolymers. *Advances in Physics: X* **9**, 2358196 (2024).
5. Alberti, S., Gladfelter, A. & Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**, 419–434 (2019).
6. Xu, Z., Wang, W., Cao, Y. & Xue, B. Liquid-liquid phase separation: Fundamental physical principles, biological implications, and applications in supramolecular materials engineering. *Supramolecular Materials* **2**, 100049 (2023).
7. Jiang, R. 1 *et al.* Emerging Implications of Phase Separation in Cancer. https://doi.org/10.1002/advs.202202855 (2022) doi:10.1002/advs.202202855.
8. Biological complexity of the phase-separated protein states. in *The Three Functional States of Proteins* 407–421 (Academic Press, 2025). doi:10.1016/B978-0-443-21809-5.00020-X.
9. Chattaraj, A. & Shakhnovich, E. I. Separation of sticker-spacer energetics governs the coalescence of metastable condensates. *Biophysical Journal* **124**, 428–439 (2025).
10. Harmon, T. S., Holehouse, A. S., Rosen, M. K. & Pappu, R. V. Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *eLife* **6**, e30294 (2017).
11. Alberti, S. & Dormann, D. Liquid–Liquid Phase Separation in Disease. *Annual Review of Genetics* **53**, 171–194 (2019).
12. Zhu, Q. *et al.* Biomolecular Condensates as Emerging Biomaterials: Functional Mechanisms and Advances in Computational and Experimental Approaches. *Advanced Materials* **37**, e10115 (2025).
13. Alberti, S., Gladfelter, A. & Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**, 419–434 (2019).
14. Mihaylova, E. M. Imaging of Live Cells by Digital Holographic Microscopy. *Photonics* **11**, 980 (2024).
15. Zorbas, C., Soenmez, A., Léger, J., De Vleeschouwer, C. & Lafontaine, D. L. Detecting material state changes in the nucleolus by label-free digital holographic microscopy. *EMBO Rep* **25**, 2786–2811 (2024).

16. Vennelakanti, V., Nazemi, A., Mehmood, R., Steeves, A. H. & Kulik, H. J. Harder, better, faster, stronger: Large-scale QM and QM/MM for predictive modeling in enzymes and proteins. *Current Opinion in Structural Biology* **72**, 9–17 (2022).

17. McManus, J. J., Charbonneau, P., Zaccarelli, E. & Asherie, N. The physics of protein self-assembly. *Current Opinion in Colloid & Interface Science* **22**, 73–79 (2016).

18. Gnan, N., Sciortino, F. & Zaccarelli, E. Patchy Particle Models to Understand Protein Phase Behavior. in *Protein Self-Assembly: Methods and Protocols* (ed. McManus, J. J.) 187–208 (Springer, New York, NY, 2019). doi:10.1007/978-1-4939-9678-0_14.

19. Saar, K. L. *et al.* Theoretical and Data-Driven Approaches for Biomolecular Condensates. *Chem. Rev.* **123**, 8988–9009 (2023).

20. Himanen, L., Geurts, A., Foster, A. S. & Rinke, P. Data-Driven Materials Science: Status, Challenges, and Perspectives. *Advanced Science* **6**, 1900808 (2019).