Thermodynamic model for the partitioning of self-replicator building blocks in complex coacervates

Pieter Brongers, *University of Groningen*

Abstract

Two essential properties that contributed to the origins of life and are necessary for the synthesis of life *de novo* are compartmentalization and self-replication. Recently a self-replicating system has been combined with complex coacervates and strong partitioning is observed. This study presents a model based on statistical thermodynamics to be able to describe this behaviour. The general principles are presented before being applied to partitioning of the replicator building block. This showed that the available space in coacervates is much lower than is expected based on its volume. It also allows to determine the free energy of partitioning. The discussion is extended to include macrocycles and linear oligomers. Only small energetic gains of a few k_bT for a macrocycle in a coacervate compared to in the solvent phase can cause major amplification in the prevalence of rings of size five and larger. The model also includes saturation effects showing that this causes larger species to become more prevalent.

Introduction

The formation of compartments is regarded to play an essential role in the emergence of life on earth.[1,2] Compartments serve to provide high concentrations of material relevant for metabolism, generate selective or specific environments and induce spatial separation necessary for a system to remain dynamic and not return to a thermodynamic equilibrium.[3,4] Besides deducing its primal form and when it appeared in the early stages of life, research is dedicated to finding suitable compartments to facilitate the synthesis of life *de novo*.

A fundamental controversy arising in the debate of the origins of life, is the definition of life itself. A common definition in the field constitutes several characteristics as summarized in figure 1.[2,3] In order to assign a system to be alive it must have information carrying molecules capable of transmitting this information to progeny by means of replication; the ability to utilize energy and resources for the formation of its own material and to stay out-of-equilibrium; and a separation from the environment to discern itself and to keep its components together. It is able to maintain its operation by staying out-of-equilibrium through dissipative processes. Mutations during replication process cause variability in the system and species that are more stable, or have a higher "fitness", will be retained. When others die, Darwinian evolution takes place. Survival of the fittest then refers to a drive to more kinetic stability.[5] Open-ended evolution can also be regarded as part of the definition of life.[6] By evolution the system will be able to acquire more functions. Both this and a combining event between two separate systems increases the complexity of the whole and can contribute to the process of going from an abiotic chemical composition, to simple forms of life, to finally the systems that we observe today.

A viable attempt to capture the above-mentioned characteristics into a molecular system is made by Otto and coworkers, the ultimate goal being that the system can be considered alive. They have found a dithiol connected to a small peptide chain with sequence GLKFK, molecule **1**, that is able to perform self-replication (fig. 2).[8] Upon oxidation disulfide bonds form and a library arises consisting of macrocycles ($\mathbf{1_n}$) of different sizes, primarily trimers, tetramers and hexamers, that inter-

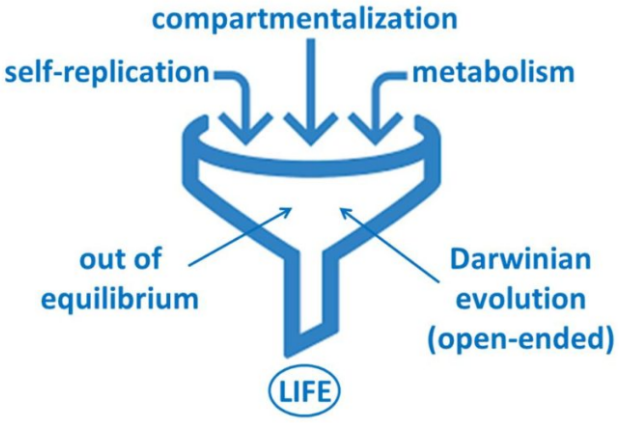

Figure 1. Life can be described as a system that exhibits self-replication, metabolism and compartmentalization. The system is self-sustaining as it keeps itself out-of-equilibrium. Darwinian evolution can occur if species of higher stability arise from mutations during replication processes while those that are less stable die. Image taken from ref. [7].

convert. At some point, cycles can stack into long fibres of tens of nm, consuming most of the available material. The cycle that does so is then called a replicator. Which cycles make up the fibres depends on the mode of agitation and the environment.[9,10]

Besides self-replicating, a stack of hexamers is able to act as a catalyst in a retro-aldol reaction and, more importantly, in the deprotection of FMOC-glycine into dibenzofulvene, the latter being able to accelerate the oxidation of monomers.[11] The second process is an example in which self-replication and metabolism is combined. By the introduction of other peptide sequences, different cycles might be competing for the same food, and perform simple evolution. The first principles thereof have also been demonstrated with two different monomers that cause two different sets of macrocycles to occur in which one is the ancestor of the other.[12]

Another objective is the incorporation of the self-replication molecule in compartments. For this purpose replicators are combined with coacervates in which they can partition. Coacervates are droplets formed by liquid-liquid phase separation in which the intermolecular interactions between components is strong enough to overcome the entropic loss of demixing.[13] Commonly this occurs using oppositely charged polyelectrolytes, i.e. ionic polymers. Polyelectrolytes can also combine into a solid phase when added to water depending on the strength of the interactions and that is not useful for dynamic incorporation of solutes. Anionic poly(acrylic acid) PAA and cationic poly(diallyldimethylammonium chloride) PDACMAC (fig. 5) form a liquid like phase when they are mixed in a charge ratio of 1.1. For much lower ratios (<0.4) they are soluble, for other ratios they form a solid complex.[14]

One of the reasons to choose for coacervates is that they do not possess a membrane, allowing a wider range of molecules to enter. It is not limited to hydrophobic solutes that can pass through a lipid membrane. Also, they are comparable to membraneless organelles that are part of real cellular systems.[16] In some theories of the origins of life, mainly by Oparin and later Lancet, coacervates have been ascribed to be the beginning of life as variaties of mixes of molecules entering the droplets might eventually lead to catalysis into new molecules or replication.[17] Coacervates have been used as membraneless cells to study rate enhancement of biomolecular reactions due to an alternate environment and high concentrations because of strong partitioning. Examples include enhanced RNA polymerization dependent on the polycation; RNA polymerization in the absence of Mg^{2+} whereas it is necessary in water; strong partitioning of nucleotides and RNA; and controllable coacervate formation and disassembly upon phosphorylation.[18–20]

Experiments have shown that **1** and the macrocycles partition strongly inside coacervates and this is promising regarding the synthesis of life.[10] To be able to compartmentalize these molecules adds complexity as there are two different phases and new opportunities such as selectively towards peptides based on the amino acids. Compartments can possibly also be exploited as a means of spatial separation of information (i.e. material) where replicators might not be able to diffuse out of the droplet

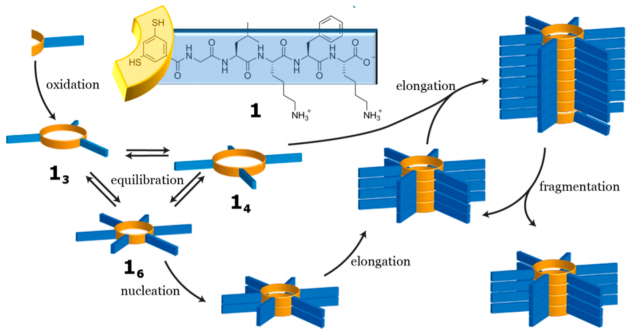

Figure 2. Self-replication system in which **1** forms rings upon oxidation, which are able to form into fibres. Image taken from ref. [15].

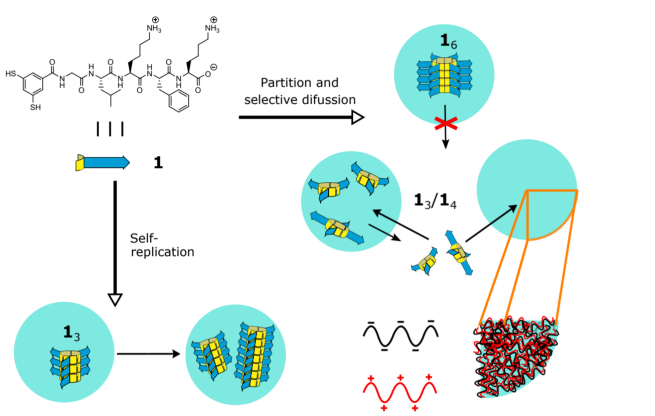

Figure 3. Possible processes involved when **1** is added to an emulsion of water and coacervates. It can diffuse into and leave the droplet and so do the macrocycles whereas replicators might not be able to. Image taken from ref. [10].

but smaller units are, and hence form the communication between cells as shown in figure 3.

In order to be able to interpret experimental data on partitioning and predict the behaviour of a network of molecules inside and outside a coacervate, a model can be made. The purpose of this study is to lay out a thermodynamic framework for the partitioning of monomers and macrocycles inside coacervates. Kinetics is not considered yet but that would be useful in order for the model to be applicable in a large dynamic assembly of all processes listed in figure 1. At first, formulations from statistical thermodynamic are introduced before being applied to the simplest case of solely monomer partitioning. Then, the discussion is extended to macrocycles. It is described how a distribution of macrocycles can be treated by an exponential function. This is applied to cycles in coacervates and water. It is shown how this can be worked out in the case that both phases are in a dilution limit. Deviations from the behaviour in the case of coacervate saturation are also explained. Moreover, linear oligomers are added and the effect of thiol oxidation on the distribution of linear and cyclic species. Finally, experiments are proposed that can test some of the observations made in this study.

Thermodynamics of partitioning

The partitioning of any species can be modelled using statistical thermodynamics. First, the principles are outlined before it is applied to the simplest case of monomer partitioning. The derivation is identical to a model on RNA adsorption on a mineral surface, coincidentally also part of origin of life research to give insight into the emergence of longer RNA chains in the presence of a mineral surface as part of the so called RNA world theory.[21] For this reason, for the first steps of the derivation, one is advised to visit [21].

Principles

For clarity and the fact that exact equations can be obtained, the coacervate is depicted as a one dimensional line that consists of *active sites*. An active site is an arbitrarily sized unit of volume from which the entire system, thus coacervate and water (solvent) is build. Useful interpretations of the size can be that of a solvent molecule, a solute molecule, or a monomer in the case that different polymers are considered that, depending on their length, take up multiple sites. For instance, in figure 4 a pink species is a tetramer that takes up four spaces. This meaning will be applied in further discussions since **1** is a monomer that can form into macrocycles of size k.

Describing a system using statistical thermodynamics relies on finding the number of possible permutations, Ω, for a certain configuration. In figure 4 the total number of species N_{tot} is the number of entities (empty, pink and green). These can permute in $(N_{tot})!$ different ways. The denominator accounts for double counting. Ω can then be used to derive thermodynamic quantities.[22] These steps are not shown here and the final result is the chemical potential of a species of size k in a coacervate, $\tilde{\mu}_k$, and is given by (eq. 1).

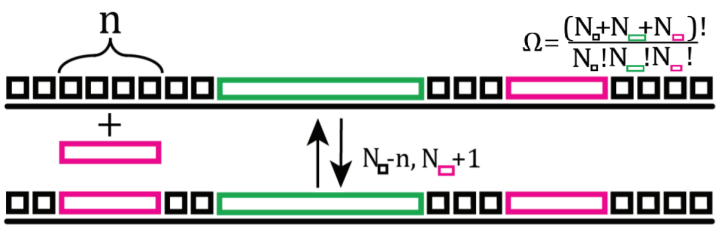

Figure 4. 1D schematic representation of a coacervate. White squares are empty sites and the pink and green depict molecules that can be adsorbed onto the sites. Image taken from ref. [21].

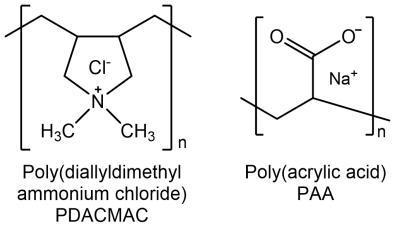

Figure 5. Oppositely charged polyelectrolytes that can form coacervates in water.

$$\tilde{\mu}_k = \tilde{\mu}_k^\circ + k_b T ln\left[\theta_k \left(\frac{\theta_W}{\theta_0}\right)^{k-1} \frac{1}{\theta_0}\right] \qquad \text{(eq. 1)}$$

In which $\tilde{\mu}_k^\circ$ is the standard chemical potentail, $\theta_k = N_k/N$, $\theta_W = W/N$ and $\theta_0 = N_0/N$ are fractions with regard to the total number of sites in the coacervate N. N_k is the number of a k-mer, $W = N_0 + \sum N_k$ is the total number of entities and $N_0 = N - \sum k \cdot N_k$ are the remaining water molecules. The same results can be obtained for the water phase where θ is replaced by χ to indicate this phase. However, since there are much more H_2O molecules than solute, this phase is very diluted and $\chi_W \approx \chi_0 \approx 1$ and the expression can be simplified to,

$$\mu_k = \mu_k^\circ + k_b T ln(\chi_k) \qquad \text{(eq. 2)}$$

Upon reaching an equilibrium in partitioning, $\tilde{\mu}_k = \mu_k$, resulting in (eq. 3).

$$\Delta\mu_k^\circ + k_b T ln\left[\frac{\theta_k}{\chi_k}\left(\frac{\theta_W}{\theta_0}\right)^{k-1} \frac{1}{\theta_0}\right] = 0 \qquad \text{(eq. 3)}$$

In here, $\Delta\mu_k^\circ$ is the standard chemical potential for partitioning and the difference between the standard chemical potential in coacervate and water. The term within the natural logarithm is an expression for the equilibrium constant of partitioning such that,

$$K_k = \frac{\theta_k}{\chi_k}\left(\frac{\theta_W}{\theta_0}\right)^{k-1} \frac{1}{\theta_0} = e^{-\beta\Delta\mu_k^\circ} \qquad \text{(eq. 4)}$$

As one can see in (eq. 4), K_k is equal to the exponential of $-\beta\Delta\mu_k^\circ$. In here $\beta = 1/k_b T$. A notable feature of this equation is that the partition constant of a species is not just the ratio of the concentration in both phases, but additional terms account for the fact that the coacervate phase can saturate. In that respect, θ_k/χ_k is the partition constant that is measured and it is equal to the true value of K_k when both phases are diluted. However, at saturation, the other terms serve as a correction factor. The apparent partition constant is $K_{k,app}$.

K_k can be expressed in terms of the number of different solutes. Considering that $\theta_W = 1 - \sum(k-1)\theta_k$ and $\theta_0 = 1 - \sum k \cdot \theta_k$,

$$K_k = \frac{\theta_k}{\chi_k} \frac{(1 - \sum(i-1)\theta_i)^{k-1}}{(1 - \sum i\theta_i)^k} \qquad \text{(eq. 5)}$$

This equation is the most general form for the equilibrium constant of partitioning of monomer and oligomers on a one dimension line.

A 1D model is limited as it does not account for the possible flexibility of larger species that is allowed in two or three dimensions. This would increase their number of configuration and decrease the entropy loss associated with a large molecule replacing multiple smaller ones. Although it is an important note, the model can still be used for qualitative predictions as the extension towards more dimensions would only shift the value of the energetic parameters (introduced later) in the standard chemical potential.[21]

Monomer only

When considering solely the monomer that occupies one site (i.e. k=1), (eq. 5) can be simplified into (eq. 6). The fractions can be expressed in terms of absolute quantities as that might be more useful when comparing it with experimental data, as discussed in an upcoming section. In here, M and N are the number of sites in water and coacervate, respectively. m and $\tilde{n}$ are the number of monomers in water and coacervate.

$$K = \frac{\theta}{\chi}\frac{1}{(1-\theta)} = \frac{\tilde{n}M}{m(N-\tilde{n})} \qquad \text{(eq. 6)}$$

The formula simplifies to an expression that can be manipulated easily. From a practical standpoint, the approximation for the dilution limit for the water phase is

not necessary. When including this, it results in,

$$K = \frac{\theta}{\chi}\frac{1-\chi}{(1-\theta)} = \frac{\tilde{n}(M-m)}{m(N-\tilde{n})} \qquad \text{(eq. 7)}$$

This partition constant stands for the reaction,

$$\mathbf{1}\,(\mathrm{aq}) + \mathrm{H_2O}\,(\mathrm{co}) \overset{\Delta\mu^\circ}{\rightleftharpoons} \mathbf{1}\,(\mathrm{co}) + \mathrm{H_2O}\,(\mathrm{aq}) \qquad \text{(eq. 8)}$$

In the case of merely monomer, it is also possible to take into consideration the relative size of **1** compared to a water molecule. Then, the meaning of a site is altered such that one site is occupied by one water molecule. A monomer is not necessarily the same size and will occupy ζ sites, resulting in the reaction,

$$\mathbf{1}\,(\mathrm{aq}) + \zeta\mathrm{H_2O}\,(\mathrm{co}) \overset{\Delta\mu^\circ}{\rightleftharpoons} \mathbf{1}\,(\mathrm{co}) + \zeta\mathrm{H_2O}\,(\mathrm{aq}) \qquad \text{(eq. 9)}$$

From (eq. 5) it follows that,

$$K = \frac{\theta(1-(\zeta-1)\theta)^{\zeta-1}}{\chi(1-\zeta\theta)^{\zeta}} \qquad \text{(eq. 10)}$$

The results of using (eq. 7) and a comparison with experimental data is shown later, as well as the results from considering ζ.

Connection with partition function

The expression for the partition constant for monomer can also be reached without considering the operations from before. For this, it is sufficient to consider Ω, the partition function Z and a factor $exp[-\beta\Delta\mu^\circ]$ that occupying a site in a coacervate is favoured over water. Then, the probability to have $\tilde{n}$ species in coacervate is given by (eq. 11).

$$p(\tilde{n}) = \frac{\Omega(\tilde{n})e^{-\Delta\mu^\circ\beta\tilde{n}}}{Z} \qquad \text{(eq. 11)}$$

For large N and M, $p(\tilde{n}) \approx p(\tilde{n}+1)$ and this yields,

$$\Omega(\tilde{n})e^{-\Delta\mu^\circ\beta\tilde{n}} = \Omega(\tilde{n}+1)e^{-\Delta\mu^\circ\beta(\tilde{n}+1)} \qquad \text{(eq. 12)}$$

From this, $K = \Omega(\tilde{n}+1)/\Omega(\tilde{n})$. Ω is given in (eq. 13).

$$\Omega(M,N,m,\tilde{n}) = \left(\frac{N!}{\tilde{n}!(N-\tilde{n})!}\right)\left(\frac{M!}{m!(M-m)!}\right) \qquad \text{(eq. 13)}$$

Solving in terms of K, and considering that $n, m >> 1$ results in (eq. 7).

However, when the size of the monomer is included, the function for Ω (eq. 14) becomes cumbersome and it not easy to solve. This is even more of a problem when multiple species are involved. The discussion from [21] provides a relatively easy model.

$$\Omega(M,N,m,\tilde{n},\zeta) = \left(\frac{(N-(\zeta-1)\tilde{n})!}{\tilde{n}!(N-\zeta\tilde{n})!}\right)\left(\frac{(M-(\zeta-1)m)!}{m!(M-\zeta m)!}\right) \qquad \text{(eq. 14)}$$

Monomer partitioning

The previously derived equations can be used to calculate the apparent partition constant as a function of the partition constant in the dilution limit, the amount of material present and the number of sites. Then, it can also be used to fit experimental data and obtain constants of the system.

Besides K, another constant can be extracted from data fitting and that is the ratio of number of sites in coacervate and water, $\phi = N/M$. The relative volume of the coacervate and water is usually known and is around 0.01-0.05. It does not necessarily mean that the active sites conform to the same ratio as the coacervate consists of polymers and the interactions between them that take up space. However, a coacervate contains water as well and if it is sufficiently swollen it is expected that the amount of water (that can be replaced) relative to the solvent is of the same order as the ratio of volumes.

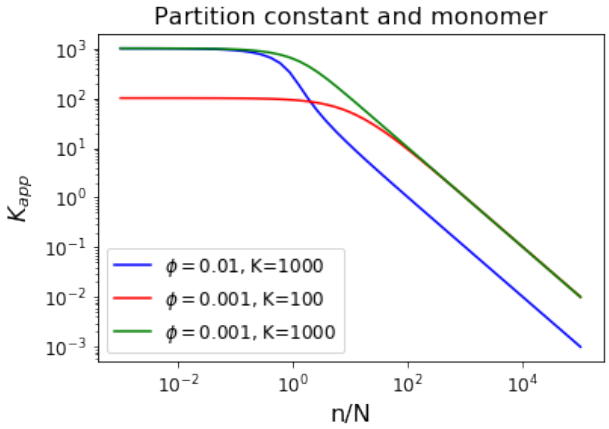

Figure 6. Using (eq. 7) the apparent partition constant can be plotted as a function of the total amount of material added compared to the number of available sites in coacervates.

Using (eq. 7) and varying K and ϕ it is possible to make a plot of K_{app} while changing the ratio of total amount of **1** added, n, to the number of sites in the coacervate. This is shown in figure 6. The most apparent feature is the decrease in K_{app} when $n \approx N$ showing that the

apparent constant decreases rapidly if more material is added than can be hosted in the coacervate, at least in the case that K is much higher than 1. In the dilution limit, K_{app} is virtually equal to K. In all three cases, the slope at saturation is equal to -1 but the height of the graph depends on ϕ. That is because a small ϕ means that there is a lot of solvent. Upon saturation, the additional material goes into the solvent but because of its volume χ stays lower than for a larger ϕ and thus K_{app} is higher.

Experimental data

Experiments have been able to measure the partition constant at various concentrations of **1** for different time intervals after addition of monomer. However, multiple issues arise in these measurements and therefore repeats (3x) at the same concentration and time show very different outcomes. The first inconvenience is that the coacervates are not constant and stable for a long period of time. What this means is that in the hours after addition of the polyelectrolytes the coacervates coalesce and break apart, so they are not properly equilibrated. After a few hours to days they will start to coalesce, form a larger phase, stick to the sides of the vial and sink to the bottom. Then they are not able to properly contribute to partitioning.

The second issue is that it is not easy to determine the concentration of material inside the coacervates. That is due to polyelectrolytes in a UPLC measurement that stick to the walls, gluing the monomers and spreading the retention time. The concentration in the coacervate is therefore determined from the content in supernatant. The experimental procedure is as follows. Polyelectrolytes are dissolved to a concentration of 50 mM in charges. Then the monomers are added at 45 °C and agitated at 1200 rpm. At different time points 0.5 ml of sample was taken and centrifuged and the supernatant was diluted such that if the total sample would be diluted the concentration is 0.2 mM, except for 0.05 mM to which nothing was added. This means dilution of 1, 2.5, 5 and 10 times for 0.2 mM, 0.5 mM, 1 mM and 2 mM, respectively. The sample is measured by UPLC and compared to a reference to determine the concentration in supernatant. The volume of coacervates can be used to determine the concentration in the coacervate. From the mass of remaining coacervates after centrifuging and the assumption of its density to be simmilar to water, the volume was estimated at 2% of the total volume.

Instead of the droplet concentration, from the concentration in supernatant a different quantity is determined, namely the ratio of contents in water and coacervate, so called the capacity factor CF. For this, the concentration in supernatant is multiplied by 0.98 (the solvent volume fraction) and subtracted from the total concentration to get the amount of monomer in the coacervate. This is then divided by the amount in supernatant ($0.98 \cdot [\mathbf{1}]_{supernatant}$). In order to get a partition constant defined as the ratio of concentration of **1** in coacervate and water, CF can be multiplied by 50. CF is directly related to the model as it is $\tilde{n}/m$. Using the definition of K_{app} in the model,

$$K_{app} = \frac{\theta}{\chi} = \frac{\tilde{n}/N}{m/M} = \frac{\tilde{n}/m}{N/M} = CF \cdot \frac{1}{\phi} \qquad \text{(eq. 15)}$$

The experimental outcome at different concentrations of monomer at 60 minutes is shown in figure 7. Since the values for CP are highly spread out, it is more informative to plot the data on a log-log plot. Moreover, then it can be compared to the theoretical models. It is indeed apparent that the trend in the data resembles that of the plots in figure 6. Data for the other time points are shown in the appendices. Some features are consistent over all graphs and those are the point of saturation and the magnitude of the capacity factor at low concentrations. However, the trends are irregular and this could have to do with the fact that the coacervates are either not equilibrated shortly after addition, or they coalesce and stick to the sides after a couple of hours. The data after one hour shows the most physically reasonable results as CP does not increase upon addition of monomer and the point of saturation is clear. Some conclusions are made but only with more reliable data these statements can be made more well-founded.

Experiment and model

Clearly, there are similarities between the model and experimental data. A clear point of saturation occurs after which the capacity factor (experiment) or apparent partition constant (model) decreases linearly on a log-log plot. Almost replicate behaviour has been observed on lysozyme partitioning in coacervates by Van Lente *et al.* when they measured the partition constant for a wide range of concentrations.[23] Finding the dilution limit and quantifying partitioning by means of a constant allows one to conclude on the partition capabilities of a

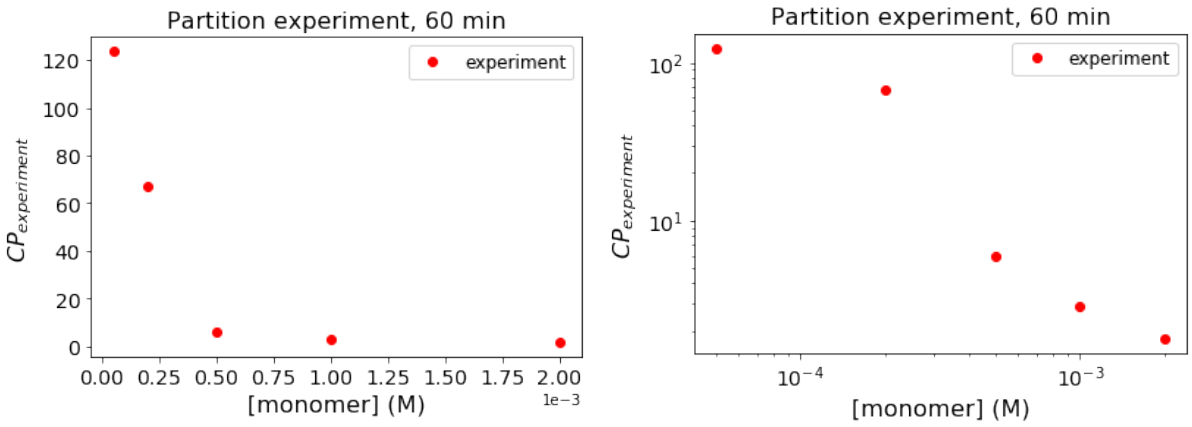

Figure 7. Experimental data on the partitioning of **1** in coacervates show saturation behaviour at 0.3 mM of monomer. This is better visualised when it is plotted on a log-log scale.

system. This is not universal in research on partitioning of (bio)molecules in coacervates as sometimes only a single concentration is taken or partitioning is observed qualitatively by turbidity measurements or spectroscopic methods.[24–27] Care should be taken on determining the partition constant when only a single concentration is examined.

The relationship between the experimental data on partitioning and the described model is discussed next.

Fit by eye

By adjusting K and ϕ in (eq. 7) a fit can be made by eye. In order to predict ϕ one should bear in mind that the point of saturation is approximately when the total amount of materials equals the total number of sites in the coacervate. That point is at $[\mathbf{1}] \approx$ 0.3 mM $\approx N$. The total amount of water sites is the concentration of water in water, i.e. 55.5 M, thus $\phi = N/M \approx 0.0003/55.5 = 5.4 \cdot 10^{-6}$. Since $CP = 124$ at dilution, from (eq. 15) it turns out that $K \approx 2.3 \cdot 10^7$. Adjusting it to $K \approx 2.7 \cdot 10^7$ results in figure 8. This means a standard binding free energy $\Delta\mu^\circ_{\mathbf{1}}$ of 17 k_bT.

The value for ϕ is much lower than $V_{co}/V_{wa} = 0.02$ indicating that the coacervate has less space for monomers to occupy than its volume suggests. This can be partly explained by the fact that the polymers take up space. Another suggesting is that it involves salt ions than are present in the coacervate. When polyelectrolytes mix and oppositely charged chains interact, many of their small counter ions are expelled to the solvent, although not all of them. It could be that the monomer only replaces the remaining ions and is not able to disrupt the interactions between polyelectrolytes even though it has charged groups. Measuring the difference in the amount of salt ions in the coacervate before and after addition of **1** could show whether this is a reasonable explanation. Then, the difference would be approximately the amount of monomer in the coacervate. Perhaps a more feasible approach is to use similar solutes that differ in their valency, for instance ADP and ATP, to see whether the capacity is related to their charge content.Insight into the ion content of a coacervate, the degree of counterions (to ensure charge neutrality), additional coions and the extent to which polyelectrolyte interactions are broken can be obtained using the theory from Schlenoff *et al.* using the Donnan equilibrium.[28]

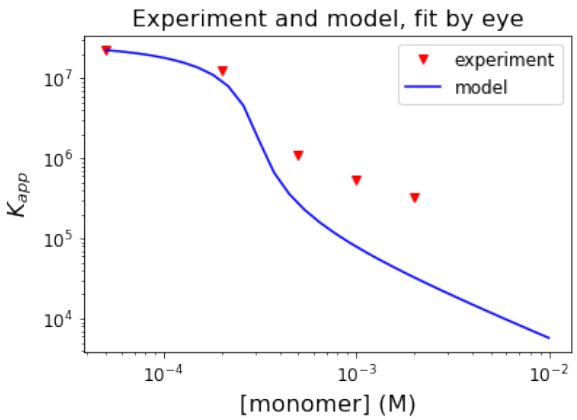

Figure 8. Experimental data is fit by eye, $K = 2.7 \cdot 10^7$ and $\phi = 5.4 \cdot 10^{-6}$.

It is interesting to see whether the small capacity originates from salt replacement or it is an intrinsic property

of this coacervate and solute system that few monomers can enter the polyelectrolyte network while not necessarily replacing ions, or a combination of both. The coacervates are dissolved in a concentration of 50 mM in charges. The saturation point of **1** is at 0.3 mM so there are about 110 polyelectrolyte charges for one charge on a monomer (containing 2+ and 1−). This would suggest that the monomer is not able to fully engage in the electrostatic interaction between polymers.

Program fit

The experimental data can also be fitted using software. For this purpose, the *Jupyter Notebook* environment is used. The *Scipy* option `curve_fit` cannot be used directly. For this purpose, the approximation that $M >> m$ is used as in (eq. 6). This equation is rewritten to free m such that m can be plotted as a function of $\tilde{n}$ with fit parameters K and ϕ.

$$K = \frac{\tilde{n}(M-m)}{m(\phi M-\tilde{n})} \approx \frac{\tilde{n}M}{m(\phi M-\tilde{n})}$$
$$\Rightarrow m = \frac{\tilde{n}M}{K(\phi M-\tilde{n})} \quad \text{(eq. 16)}$$

The experimental data is rewritten in terms of m and $\tilde{n}$ using (eq. 17).

$$CP = \frac{\tilde{n}}{m} = \frac{\tilde{n}}{n-\tilde{n}} = \frac{n-m}{m}$$
$$\Rightarrow \tilde{n} = \frac{n\cdot CP}{1+CP} \;\; and \;\; m = \frac{n}{1+CP} \quad \text{(eq. 17)}$$

The fit parameters from this method are $\phi = 3.47\cdot 10^{-5}$ and $K = 1.51\cdot 10^{5}$. This differs by one or two orders of magnitude from the fit by eye. The effect of the fit parameters is shown in figure 10 and the fit is not good at dilution. This has partly to do with the fact that only for saturation $\tilde{n}$ and m are substantial and hence fit parameters will attempt to minimize deviations from that region in the $m,\tilde{n}$-plot. Moreover, it is because of the increase in ϕ which causes the point of saturation to move to 1.8 mM whereas clearly a form of saturation should occur at 0.3 mM. This is even more clear in figure 9. It shows that the maximum value for $\tilde{n}$ approaches 1.8 mM, higher than expected from figure 7. This would mean that the maximum capacity of the coacervate is higher. Possibly the coacervate swells slightly upon saturation, increasing ϕ and lowering K_{app} in (eq. 15). Coacervates swell upon increasing salt concentration until the polyelectrolyte interactions are completely screened by ions and it dissolves.[29,30] Uptake of monomers could increase the ionic strength in the coacervate, causing it to swell and be able to accommodate more species. Zacharia and coworkers have found a system in which solute partitioning grows when its total concentration increases as the molecules improves the hydrophobicity in the coacervate.[31,32] There are no indications that this is the case as K_{app} does not increase upon higher monomer concentrations. Yet, it does for some of the other time points but this has more likely to do with experimental errors. However, it is a useful concept to consider when trying to explain the partitioning behaviour of **1**.

Another reason could have to do with oxidation of monomers inside the coacervate. Previous experiments has shown that in the presence of pDACMAC accelerated oxidation takes place whereas oxidation from air usually takes a couple days.[10,33] If some of the monomer inside the coacervate is converted to other species it will decrease the concentration of monomer in supernatant. From the method to determine the coacervate content this would imply a value that is too high. Then, CP is also increased and the data points end up higher than what the model suggests.

Monomer size

In order to account for the relative size of **1** to H_2O, a ζ parameter has been introduced in (eq. 10). ζ is a measure of the effective volume that a monomer takes up in the coacervate in comparison with a water molecule and as such can give insight into the environment of the molecules. If it is higher than what one would expect based on a crude calculation of the volume of the molecule, the monomer replaces more water molecules and it is not able to be as tightly incorporated within a coacervate as water is.

Writing (eq. 10) in terms of absolute values:

$$K = \frac{\tilde{n}(1-(\zeta-1)\frac{\tilde{n}}{\phi M})^{\zeta-1}}{m\phi(1-\zeta\frac{\tilde{n}}{\phi M})^{\zeta}} \quad \text{(eq. 18)}$$

Now, saturation occurs when $\zeta\tilde{n}/(\phi M) \to 1$, where $\tilde{n} \approx n \approx N$. This is similar as before apart from the addition of ζ, hence ϕ will be increased by a factor ζ and K decreases by that same amount. Considering that ζ is in the order of ten to hundred, K will lower but it will still be multiple orders of magnitude high.

What value could one expect for ζ? ζ is an indication for the size of **1** relative to a water molecule. The molecular mass of **1** is 760.982 g/mol and for water it is 18.015

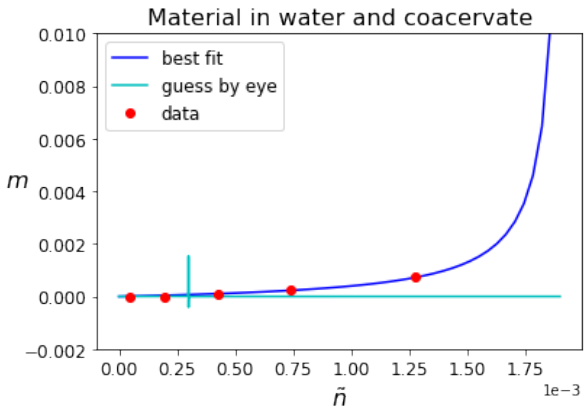

Figure 9. $m,\tilde{n}$ plot that is used to find fit parameters.

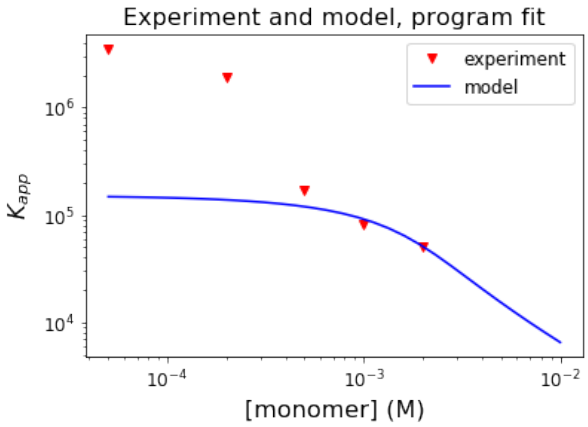

Figure 10. Experimental data fitted using software, $K = 1.51 \cdot 10^5$ and $\phi = 3.47 \cdot 10^{-5}$.

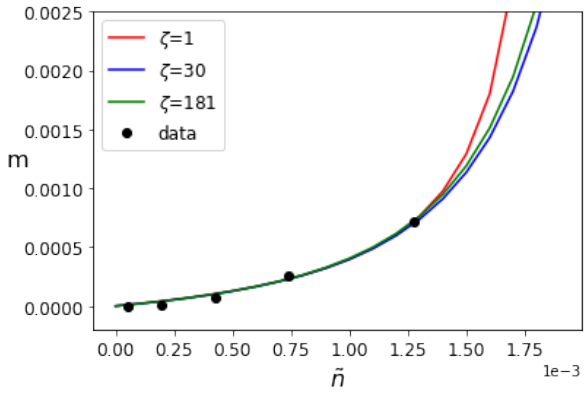

Figure 11. While considering the size of **1** might reveal more information about the system, it turns out that different values for ζ results in very similar results.

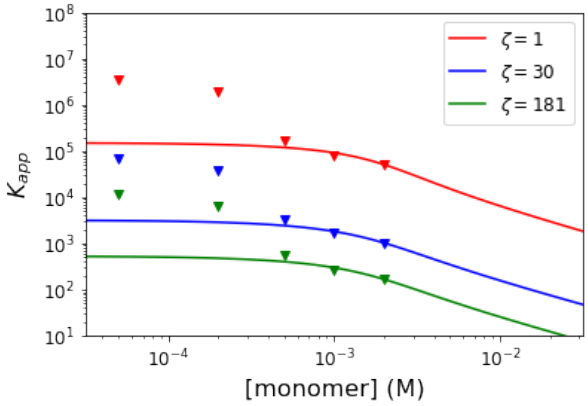

Figure 12. Different sets of parameters give equivalent results. — $\zeta = 1$, $\phi = 3.47 \cdot 10^{-5}$, $K = 151380$; — $\zeta = 30$, $\phi = 1.75 \cdot 10^{-3}$, $K = 3204$; — $\zeta = 181$, $\phi = 0.0105$, $K = 534$.

g/mol. Then, ζ is expected to be around 42. A crude estimation of ζ can also be made based on the dimensions of a stack of hexamers. One layer is roughly 0.5 nm thick with a diameter of 3.2 nm, thus the volume of one molecule is $(0.5 \cdot 1.6^2\pi)/6 = 0.67$ nm^3.[15] The volume of H_2O is 0.030 nm^3, hence $\zeta \approx 22$. When $\mathbf{1_6}$ is aggregated the volume of the individual molecules is probably reduced due to favourable interactions and ζ should be slightly higher. In any case, the range of ζ is quite large but the value from the model is expected to be in this range.

The same methods for data fitting can be applied, i.e. expressing (eq. 18) in terms of m. From figure 11 and 12 it is clear that different sets of data give similar results. That is not unexpected considering that ζ, ϕ and K are proportionally related. The major difference is that, due to a power of ζ, the plot is more sensitive to small changes of $\tilde{n}$ and hence the plots do not perfectly overlay. However, in order to determine what ζ fits best and to get a precision of about 5 (relevant for interpreting the size), more repeats of the same experiment with more accurate concentration measurements are necessary.

The issue of a discrepancy in the capacity of the coacervate still remains as the principles of the model, that is constant number of sites and size of sites, do not change. In order to account for a changing environment in the coacervate, the model should be extended, for instance by including sites that become accessible when a certain level of saturation is reached. This would add a level of complexity to the derivation from the beginning and this is not further examined.

Macrocycle partitioning

In the upcoming sections the discussion of partitioning is extended to more molecules such as macrocycles and linear oligomers of monomer **1**. The reactions that play a role are introduced. The thermodynamic model outlined at the start is expanded to be able to describe the multicomponent system. Length dependent affine functions for the standard chemical potential enable to express equilibrium constants in terms of energetic parameters. From this it follows that species are exponentially distributed. This is applied to a system containing solely cyclic species and a combination of cyclic and linear species. Finally, the effect of saturation is discussed.

Reactions

When considering more than one species, it is important to take into account the reactions that can occur as ultimately these decide what products are formed and in what amounts. The reactions in a monomer-only system are limited as it can only perform a transfer from one phase to the other. Including macrocycles complicates the system since beside partitioning, they can react with one another, both in water and coacervate. In this section it is discussed what these reactions are and the implications of describing them in one way or another.

There are two new reactions that come into play. Those are thiol disulfide exchange and oxidation of thiols. A thiol disulfide exchange reaction converts a thiol and disulfide into two of the same kind while connected groups are exchanged, i.e. $R_1SH + R_2SSR_3 \rightleftharpoons R_2SH + R_1SSR_3$. In case of a thiol and a single type of chain (R=X−GLKFK) this would not have an effect. However, **1** is a dithiol that forms into rings. If this reacts with one of the thiols of a linear species, for example a monomer or dimer, then the ring is broken up. Since both the new thiol and disulfide are connected through the remainder of the original ring, one linear species is formed.

This linear molecule is reactive and can combine with another ring to form an even longer oligomer. It is also able to perform a backbiting reaction in which a terminal thiol reacts intramolecularly with a disulfide moiety, releasing a linear molecule. The net reaction over two steps is $[m] + [n]^* \rightleftharpoons [m-k] + [n+k]^*$, in which $[m]$ is a ring and $[n]^*$ is linear. Both reactions are still examples of thiol disulfide exchange. They are relevant as they allow to construct a mechanism for the conversion of one macrocycle into another with only catalytic contributions from linear species. For instance in the case of trimer and tetramer conversion,

$$\text{Net reaction: } 4\,[3] \rightleftharpoons 3\,[4]$$

$$[3] + [1]^* \rightleftharpoons [4]^* \overset{+[3]}{\rightleftharpoons} [7]^* \rightleftharpoons [4] + [3]^* \overset{+[3]}{\rightleftharpoons}$$

$$[6]^* \rightleftharpoons [4] + [2]^* \overset{+[3]}{\rightleftharpoons} [5]^* \rightleftharpoons [4] + [1]^*$$

Any reaction of the form $n\,[m] \rightleftharpoons m\,[n]$ can be constructed in this way. Using this method macrocycle conversion can also be written as $[m] + [n] \rightleftharpoons [m+n]$. This requires less steps and combining multiple of these reactions adds up to the former reaction. The latter form is used in the upcoming discussions of thermodynamics as it entails more general features such as a constant form and magnitude of the associated equilibrium constant. These still play a part if the reaction is written as $l\,[k] \rightleftharpoons k\,[l]$ but since it can be written as a combination of multiple reactions, the equilibrium constants are not always equal.

Two linear species can also react by thiol disulfide exchange. An overview of all reactions and partitioning is shown in the first column of table 1. In here, an asterisk ($*$) signifies a linear species and a tilde ($\sim$) the fact that the species is in a coacervate.

As mentioned before, oxidation of two thiols into a disulfide bond occurs as well. This reaction is first order in thiol with a reaction rate constant $k = 2 \cdot 10^{-4}$ min^{-1}.[33] Thiol disulfide exchange has a rate constant of $k = 6.48 \cdot 10^6$ $\text{M}^{-1}\text{min}^{-1}$ and is first order in both thiol and disulfide.[33] The rate of oxidation is much lower and for this reason it is assumed that all macrocycles and linear species are in equilibrium.

Extension on thermodynamics

While expanding on the number of distinct species will increase complexity, the thermodynamic framework laid down at the beginning still holds. The formulae for the chemical potential of species and the partitioning constant that was derived from it, have already been explained in the context of monomers and polymers. Now, these are used as the are presented instead of reducing them down by only looking at monomer partitioning. It only requires small additions to give meaning into the standard chemical potential. Then, those can be used to derive expressions for the equilibrium constants for partitioning and exchange reactions.

$[m]+[n] \rightleftharpoons [m+n]$	K	$\frac{\chi_{m+n}}{\chi_m \chi_n}$	$e^{\beta a}$
$[\tilde{m}]+[\tilde{n}] \rightleftharpoons [\tilde{m}+\tilde{n}]$	$\tilde{K}$	$\frac{\theta_{m+n}}{\theta_m \theta_n}\theta_W$	$e^{\beta \tilde{a}}$
$[m]^*+[n]^* \rightleftharpoons [m-k]^*+[n+k]^*$	$K_{ex,l}$	$\frac{\chi^*_{m-k}\chi^*_{n+k}}{\chi^*_m \chi^*_n}$	$e^{-\beta \cdot 0}=1$
$[m]+[n]^* \rightleftharpoons [m+n]^*$	$K_{ex,m}$	$\frac{\chi^*_{m+n}}{\chi_m \chi^*_n}$	$e^{\beta a}\cdot e^{-\beta(b^*-b)m}$
$[\tilde{m}]^*+[\tilde{n}]^* \rightleftharpoons [\tilde{m}-\tilde{k}]^*+[\tilde{n}+\tilde{k}]^*$	$\tilde{K}_{ex,l}$	$\frac{\theta^*_{m-k}\theta^*_{n+k}}{\theta^*_m \theta^*_n}$	$e^{-\beta \cdot 0}=1$
$[\tilde{m}]+[\tilde{n}]^* \rightleftharpoons [\tilde{m}+\tilde{n}]^*$	$\tilde{K}_{ex,m}$	$\frac{\theta^*_{m+n}}{\theta_m \theta^*_n}\theta_W$	$e^{\beta \tilde{a}}\cdot e^{-\beta(\tilde{b}^*-\tilde{b})m}$
$[m] \rightleftharpoons [\tilde{m}]$	K_m	$\frac{\theta_m}{\chi_m}\left(\frac{\theta_W}{\theta_0}\right)^{k-1}\frac{1}{\theta_0}$	$e^{-\beta((\tilde{a}-a)+(\tilde{b}-b)m)}$
$[m]^* \rightleftharpoons [\tilde{m}]^*$	K^*_m	$\frac{\theta^*_m}{\chi^*_m}\left(\frac{\theta_W}{\theta_0}\right)^{k-1}\frac{1}{\theta_0}$	$e^{-\beta((\tilde{a}^*-a^*)+(\tilde{b}^*-b^*)m)}$

Table 1. Overview of relevant reaction with their equilibrium constants being expressed in terms of species fractions and standard chemical potential parameters. The terms in blue can be discarded in a diluted coacervate phase.

The standard chemical potential of a macrocycle in coacervate is described by an affine function as most thermodynamic contributions will be either constant or linearly dependent on the length.[34] Examples are the interaction energy of peptide chains or with the coacervate, and the free energy of formation. This yields,

$$\tilde{\mu}^\circ_k = \tilde{a} + \tilde{b}\cdot k \qquad \text{(eq. 19)}$$

In here, $\tilde{b}$ is assumed to be negative arising from favourable interactions that accumulate with length such that macrocycles have a driving factor for formation of larger species. The same reasoning applies to the water phase and to linear species of various lengths. Thus for the chemical potential of molecules,

$$\begin{aligned}
\text{Cyclic, water; } \mu^\circ_k &= a + b\cdot k \\
\text{Linear, water; } \mu^{\circ *}_k &= a^* + b^*\cdot k \\
\text{Cyclic, coacervate; } \tilde{\mu}^\circ_k &= \tilde{a} + \tilde{b}\cdot k \\
\text{Linear, coacervate; } \tilde{\mu}^{\circ *}_k &= \tilde{a}^* + \tilde{b}^*\cdot k
\end{aligned}$$

Ultimately the energetic parameters will decide on the magnitude of the equilibrium constant for partitioning and disulfide exchange. For any reaction, $K = exp[-\beta\Delta\mu^\circ]$, where $\Delta\mu^\circ$ is a function of the standard chemical potential of the species involved. The difference in the standard chemical potential will be the sum of the standard chemical potential of reactants and products multiplied by their stoichiometry, with a negative sign for the first.

As an example, the partition reaction of a macrocycle of size k from water to a coacervate is in equilibrium when $\mu_k = \tilde{\mu}_k$. From (eq. 4) one can recall that,

$$K_k = \frac{\theta_k}{\chi_k}\left(\frac{\theta_W}{\theta_0}\right)^{k-1}\frac{1}{\theta_0} = e^{-\beta\Delta\mu^\circ_k} \qquad \text{(eq. 20)}$$

With $\Delta\mu^\circ_k = \tilde{\mu}^\circ_k - \mu^\circ_k$. One can compute for this delta such that $\Delta\mu^\circ_k = (\tilde{a}+\tilde{b}\cdot k)-(a+b\cdot k) = (\tilde{a}-a)+(\tilde{b}-b)k$. The same method can be applied to an exchange reactions of the form $[m]+[n] \rightleftharpoons [m+n]$. Equilibrium means that $\mu_{m+n} = \mu_m + \mu_n$. Then, $\Delta\mu^\circ = (a+b\cdot(m+n))-(a+bm+a+bn) = -a$. An overview of equilibrium constants for the previously discussed reactions is given in table 1. The origins of $\Delta\mu^\circ$ for partitioning can be plentiful as summarized by Nakashima *et al.*[16] Decreased polarity in coacervates can favour hydrophobic solutes. Charged molecules can interact with polyelectrolytes. These can also replace small ions meaning that entropic gains highly contribute to the free energy release. Specific hydrogen bonding pairs can play a part as well. A factor that disfavours uptake of large solutes is that the mesh of the polyelectrolyte network has to be deformed. This could contribute to the explanation that $\mathbf{1_3}$ fibres are formed rather than $\mathbf{1_6}$ stacks, whereas in aqueous solution the latter will form. Though this phenomenon is still not understood as other factors can play a role such as fast oxidation by PDACMAC or the impurities in PDACMAC, consuming all linear species and disallowing exchange reactions, hence preventing hexamer formation in the first place.[10]

In going from a 1D description to multiple dimensions, it appears that the parameters in the function of the standard chemical potential shift.[34] The calculations have not been repeated while including a factor that takes into

account the increased degree of freedom for oligomers to fold. Therefore, the magnitude of the translation of the parameters is not known.

Exponential distribution

Assuming equilibrium over oligomers that can interconvert via the reaction [m] + [n] $\rightleftharpoons$ [m + n] enables to write the distribution of species as an exponential function of form $\chi_l = AB^{l-l_{min}}$, where l_{min} is the minimum length of a species and A and B are constants.[34,36] This is the only function that satisfies the conversation of number of monomers in the reaction. As an example, let us look at the macrocycle distribution in water for which $\chi_l = AB^{l-3}$. $l_{min} = 3$ as cyclic dimers are not possible. At equilibrium,

$$K = \frac{\chi_{m+n}}{\chi_m \chi_n} \qquad \text{(eq. 21)}$$

The exponential distribution will results in,

$$K = \frac{AB^{m+n-3}}{AB^{m-3}AB^{n-3}} = \frac{B^3}{A} \qquad \text{(eq. 22)}$$

The length contribution cancels due to the fact that exponents add up when multiplied. This means that K does not have a length dependence and that should be the case for this reaction (table 1). A linear function $\chi_k = A + B \cdot k$ would not yield the same result as the numerator and denominator contain either a sum of lengths $B(m+n)$ or a product and sum $(ABm + ABn + B^2mn)$. Regarding that K is equal for all reactions, this would mean that A or B has a length dependence and this cannot be true as they are constants. Another reason to believe that the exponential form is the correct description comes from the fact that if it were not and another function would fit the conversion criterion, then the system has descriptions for multiple states for thermodynamic equilibrium and this cannot occur.

For a diluted medium the parameters in the distribution can be found algebraically. For the smallest species, $l = l_{min}$ hence $\chi_{l_{min}} = AB^0 = A$. From (eq. 22) it is apparent that $B = (AK)^{1/3}$. The macrocycle distribution in water is therefore described by $\chi_l = \chi_3((\chi_3 K)^{1/3})^{l-3} = K^{-1}(\chi_3 K)^{l/3}$. For any exponential, A is the fraction of the smallest species in the series and as such is an indication for the content of the system. B is the ratio of two consecutive species χ_{l+1}/χ_l and is a measure for the prevalence of larger species. B is always smaller than 1 to ensure that fractions are convergent to zero for large l. The magnitude gives insight into how fast zero is approached. An example on what a distribution looks like is shown in figure 13. Figure 14 displays a mixture of macrocycles of different sizes made from a monomer that is similar to **1**. Although the distribution is not perfectly exponential for small cycles, the experiment shows signs for the use the above-described function. An explanation could be that for rings larger than seven, the addition of a monomer contributes purely linearly to the energy. For smaller cycles, specific configurations could occur in one that are not possible for others, thus decreasing its energy disproportionally.

The exponential function applies to cyclic and linear species, in both phases. For linear molecules the functions are $\chi_l^* = CD^{l-1}$ and $\theta_l^* = \tilde{C}\tilde{D}^{l-1}$, for water and coacervate, respectively. It also applies when saturation

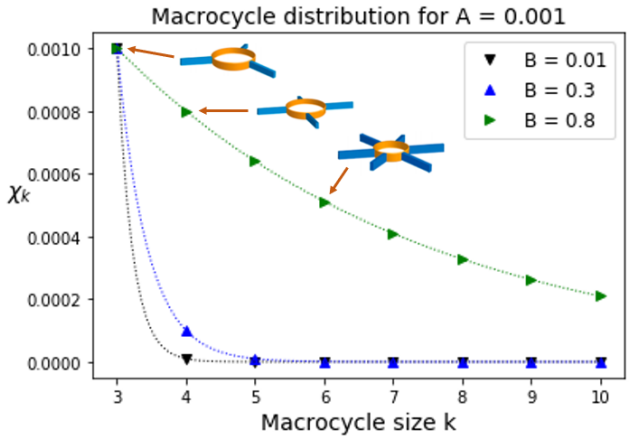

Figure 13. Visualization of a distribution as follows from the fact that the macrocycle distribution is described by an exponential function.

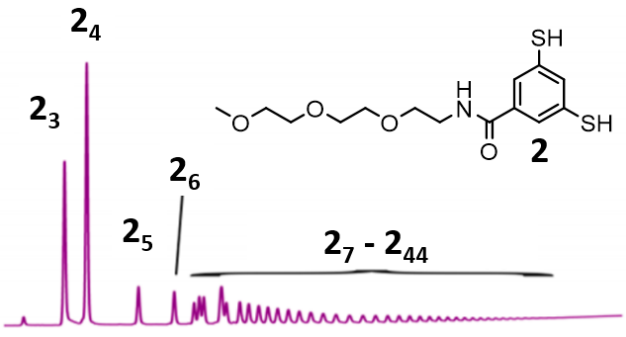

Figure 14. Exponential distribution of macrocycles derived from a similar monomer. Image taken from ref. [35].

effects are considered as again the length dependence cancels. However the previous derivation for obtaining B, and from there an exact expression, does not work since the equilibrium constant includes θ_W. The details and how to use the distribution function are discussed later and in the appendices.

A final note on the use of exponential functions is that it allows one to find expressions for the total mass and total number of species, quantities that are relevant for finding θ_W and θ_0 and working with linear oligomers. Those involve infinite sums over all fractions. These sums have exact solution and are shown in (eq. 23) and (eq. 24). These are applied throughout the subsequent sections.

$$\sum_{l=k}^{\infty} x^l = \frac{x^k}{1-x} \qquad \text{(eq. 23)}$$

$$\sum_{l=k}^{\infty} l \cdot x^l = \frac{x^k(k-(k-1)x)}{(x-1)^2} \qquad \text{(eq. 24)}$$

Application of the model

The exponential distribution can be used to gain insight into the distribution of species in the solvent and coacervate. Especially the parameters B, D, $\tilde{B}$ and $\tilde{D}$ since these are the ratio of two consecutive species with size. One could attempt to solve analytically for certain quantities (distribution as function of total material, $B/\tilde{B}$ or others depending on the purpose) however this is very limited. However it is possible to obtain results from the above discussed thermodynamics and this is shown for some scenarios but first the limitations are elaborated.

When one would like to solve for the fraction of a trimer, and then imply the other fractions, as a function of the total mass, (eq. 24) is used. Even for a single phase and one type of species this is algebraically not solvable since $k=3$ and $x=B=(\chi_3 K)^{1/3}$. If a second phase or linear species are included the sum of terms complicates it even more. In order to work around this, all exponential parameters are expressed in terms of one of them and this is then solved. For this B is a natural reference point as it is the distribution of macrocycles in the solvent phase. From this the other parameters can be derived. A detailed description is shown in the appendices.

This method is only applicable when all phases are in a dilution limit. When this is not the case, the equilibrium conditions for interconversion and partitioning contain additional terms accounting for the limited number of configurations (table 1). θ_W and θ_0 are dependent on all species in the mixture. In principle, what this means is that in order to know the ratio of two species one must know them in advance to solve for θ_W and θ_0. One could express them in terms of distribution parameters but any attempt to manipulate those equations, even for qualitative phenomena only, did not succeed. When the model will be expanded to more than one building block, and so introducing more species that are dependent on one another, it is essential to find different methods or new merits to describe a system while considering saturation effects. Perhaps similar effects are described in the literature but this has not been investigated yet. For now to work around this problem, only one phase is considered in which saturation occurs.

For a distribution of macrocycles in water and coacervate, $\tilde{B}/B = e^{-\beta(\tilde{b}-b)}$. $\tilde{b}-b$ is the extent to which the linear contribution to the standard chemical potential of a macrocycle is more favourable in the coacervate compared to water. It is expected that the length dependence in a coacervate is stronger as the less polar environment allows the cycle to fold more freely as the hydrophobic parts of the molecule do not have to be screened. Also, the rings can release charge free energy upon partitioning, a contribution that increases with size. Even when this difference is only 1 k_bT to 3 k_bT, $\tilde{B}/B$ ranges from 3 to 20 and thus larger species become more prevalent in the coacervate compared to water. Another way to view it, is by considering how the fraction of a cycle relative to the total number increases in going from water to coacervate, i.e.

$$Q_l = \frac{\theta_l/(\sum_{i=3}^{\infty}\theta_i)}{\chi_l/(\sum_{i=3}^{\infty}\chi_i)} \qquad \text{(eq. 25)}$$

Figure 15 shows Q_l as a function of $\tilde{b}-b$ and the ring size. Even for differences for b of only 2 or 3 k_bT, the presence of a macrocycle in a coacervate relative to solvent can increase with 1 to 3 orders of magnitude.

Next, it is interesting to be able to include linear species as they might contribute to fibre growth. This growth is not completely understood. A proposed mechanism is that the a macrocycle of the same size as the fibre diffused onto one of the ends. It is also possible that the ends act as a template onto which smaller linear and cyclic species diffuse. Lastly, adsorption of species on the sides of the fibres can concentrate them before they diffuse to the ends. If linear species turn out to be relevant for the mechanism, it is useful to include them in

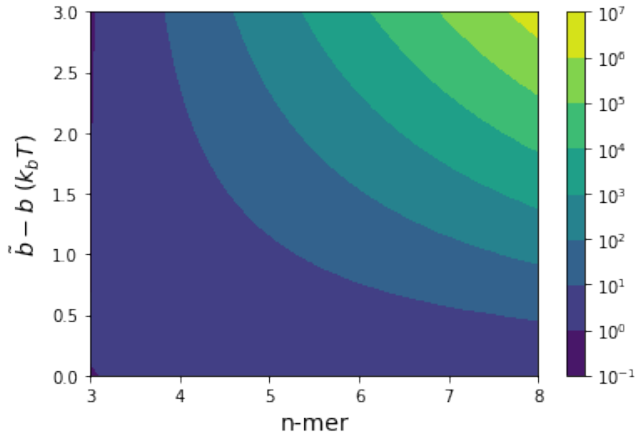

Figure 15. For small difference in the linear component of the standard chemical potential, Q_l can become several order of magnitude.

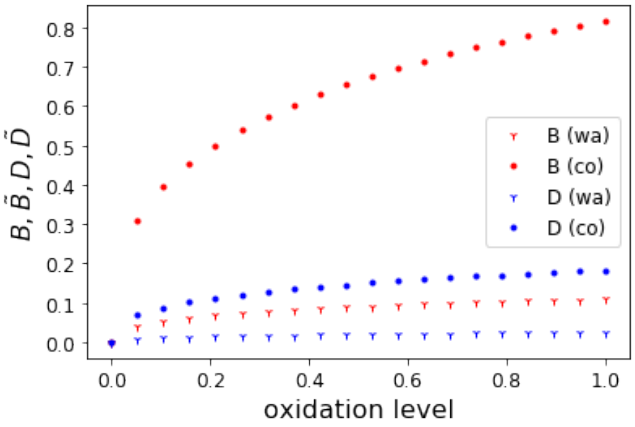

Figure 16. As the oxidation level increases linear oligomers are converted to ring structures and the distribution is pushed towards larger species. For model parameters see text.

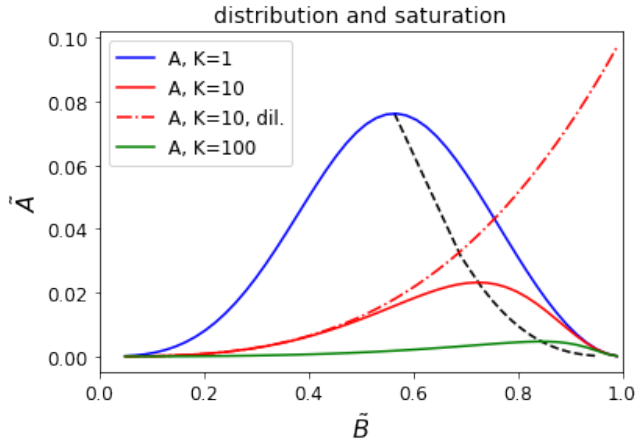

Figure 17. In a dilution limit $\tilde{B}$ increases with larger $\tilde{A}$ such that a distribution is moved towards larger cycles for more total mass. A saturated phase experiences a maximum in $\tilde{A}$ limiting the amount of trimer. In here, $K = \tilde{K}$.

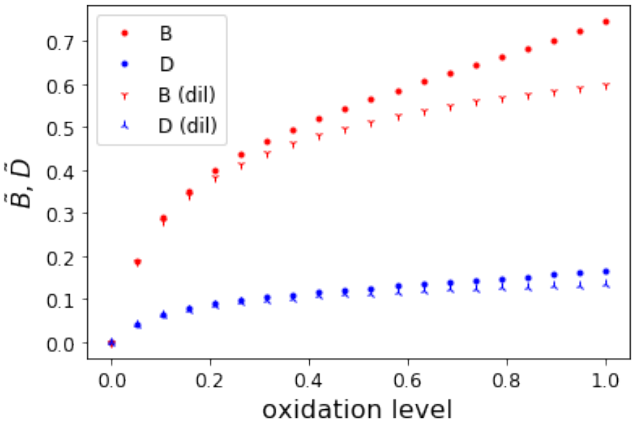

Figure 18. Saturation behaviour causes the distribution to be pushed towards larger as the oxidation level increases. In here $\tilde{K} = 2.7$ and the **1** eq. to the number of sites is 0.9.

the model. For instance, if linear oligomers attaches to the end of a fibre in the template mechanism, the longer it is, the stronger the interaction and the less likely it is to detach. Also, a linear species is more flexible that a cycle to fit on the fibre end, especially if the cycle is not of the same size as those of the fibre.

Figure 16 shows the exponential parameters as a function of the oxidation level. The oxidation level is the number of disulfide bond over the sum of disulfide bonds and dithiols. It is apparent that upon oxidation of the system, the distribution is pushed towards larger species. At low oxidation level this increases more quickly before becoming roughly linear. For this set of parameters, it is clear that for energetically unfavourable species (water phase or linear) the distribution is mainly concentrated towards small molecules (small B).

The input parameters are as follows. The total mass is 1 mM monomer equivalent in a water phase of 1.39 M. The water concentration is lowered by 40 times as a rough estimate of the size difference between **1** and H_2O, in principle considering 40 H_2O molecules as one. $\phi = 10^{-5}$ since that is the order of magnitude derived from monomer partitioning. The constant term of the standard chemical potential for cyclic and linear is 2 k_bT in water and 1 k_bT in coacervate. This is relatively small reflecting the fact that most energetic effects will scale with length and changes by 1 or 2 k_bT do not change the results. More importantly are the length dependent terms. These can be changed according to experiment data or to compare building blocks. For now, b is set to

2.5 k_bT, roughly equal to the binding energy of two peptide chains.[15] In a coavervate this is more favourable by 2 k_bT. For linear species $b^* = 1$ k_bT based on that peptide interactions constrain the flexibility of the molecule more than for a cycle.

In order to get insight into the energetics and to the extent that the model resembles experiments, measurements of species in water and coacervate can be done. UPLC allows to determine the concentration of all species, to the extent that they are present in sufficient amounts. From the degree of cyclic species and linear oligomers, the oxidation level is apparent and the ratio of consecutive species can be fitted. If only monomer and cyclic trimer can be measured (i.e. A, C, $\tilde{A}$ and $\tilde{C}$) it is also possible to study their behaviour as a function of oxidation level.

From figure 16 it is apparent that for the water phase only monomers and cyclic trimers are present. This could hinder the formation of hexamer stacks that commonly form.[8,15] In the absence of coacervates, other cycles and linear oligomers would be present as well (see appendices), potentially enabling the nucleation of hexamers into a stack of two before growing into large fibres. In the presence of coacervates only small species are present and it is highly unlikely for two hexamers to nucleate. Even if if occurs new building blocks for fibre growth are small making templating and adsorption on the sides less strong. The likelihood for a new hexamer to stack is even smaller.

Lastly, saturation effects are considered. For a diluted phase $\tilde{B}$ increases with $\tilde{A}$, effectively saying that when the total mass increases, the distribution goes to larger species. However, for a phase subject to saturation, this cannot occur as at a certain point all sites are occupied. Hence, $\tilde{A}$ is limited by the maximum fraction that the trimer can have, also visible in figure 17. It is expected that for $\tilde{B}$ at $\tilde{A}_{max}$, $\theta_0 = 0$ as this is the number of remaining solvent molecules. From the figure in the appendices, it is clear that this is not the case. $\theta_0 = 0$ at larger $\tilde{B}$ suggesting that for some $\tilde{A}$ two values for $\tilde{B}$ are possible. The nature of this is not further looked into. This could limit a measurement of $\tilde{A}$ at saturation as it is not known whether the maximum is obtained or the one for $\tilde{B}$ at which $\theta_0 = 0$. However, this different is not very large so from a measument of $\tilde{A}_{saturation}$, if $\tilde{B}$ is known, $\tilde{K}$ can be determined or vice versa.

Saturation effects pushes the distribution to larger species. From figure 17 it is apparent that for a certain $\tilde{A}$, $\tilde{B}$ is higher compared to dilution. This effect is stronger close to saturation. This can also be deduced from the equilibrium condition in table 1. For a cyclic exchange reactions, θ_W is included. This is between zero and one. Upon saturation it decreases and thus the fraction of larger species has to increase relative to the smaller ones to retain a constant $\tilde{K}$. For a system containing cyclic and linear species, the fact that saturation favours larger species is also the case as visualised in figure 18.

Experiments

As a matter of validation for the model and to examine its applicability in the construction of coacervate systems, experiments can be conducted. Some ideas have been mentioned throughout the text and listed here.

One of the experiments can be concerned with the partitioning of monomer to gain insight into the small capacity of coacervates. The hypothesis is that the monomers are solely capable of replacing salt ions. To test this, the partitioning of the nucleotide sequence AMP, ADP and ATP can be measured. This should be done at similar concentration as for **1** (0.05-2 mM) since the structure of the molecules are similar with an aromatic end and ionic tail. As the valency increases linearly the maximum molar occupancy of the coacervates should decrease in this manner. Also, for these molecules again an affine function for the standard chemical potential can be proposed. The constant contribution is related to the adenosine while the length dependence originates from the phosphate groups. The partition constant can be obtained from partition measurements, similarly as to **1**. As the molecules do not have side reactions such as oxidation, it is reliable to measure the supernatant concentration and imply the coacervate concentration. Then the partition constants should fit to the expression in table 1.

Solutes that are closer to the current system would be those in which one or two lysines are replaced with a neutral amino acid such as serine or glutamine. By changing one lysine, the molecule is neutral although it still possesses charges. Then it can become apparent whether the overall charge or the individual charges play a role. If it is the total charge, then virtually no salt ions will be replaced, limiting the total capacity. For the second case, the number of charges is less, so less charge will be replaced and thus increasing the capacity. The entropy

gain upon partitioning decreases as less salt ions are released to the solvent, lowering the partition constant. A completely non ionic peptide can be used as well to examine the partitioning if no charges and little polarity is involved. It might still partition strongly due to the less polar environment of the coacervate. This gives a measure for the hydrophobic contribution to the free energy of partitioning. A more feasible soluble hydrophobic molecule such as a nucleobase can be used but its nature and size is different from peptides that it only gives an indication of partitioning of a hydrophobic molecule.

Furthermore, a careful investigation for the exponential distribution would be useful in order to see whether it makes sense to use the form in future discussions. Its validity is based upon two principles. The first is that macrocycle interconversion is sufficiently faster than oxidation so that they virtually equilibrate. The second is that the standard chemical potential is linearly dependent on the size. Any deviations from this would suggest that other processes are involved such as specific favourable folding. Another process that is recently observed is the formation of aggregates consisting of cycles of various sizes. This limits the use of experimental data for the purpose of determining the distribution as the measurements will also contain aggregate cycles that are not involved in interconversion reactions. Aggregates are observed at least for at concentrations of 500 μM of monomer.

For this reason it is advised to do measurements at a concentration of monomer equivalent as low as possible. Since saturation occurs at 0.3 mM, a good starting point would be 0.1 mM. Both the composition in water and coacervate should be monitored. The model suggests that for the water phase almost only monomer and cyclic trimer are present (fig. 16). These can still be used to fit but are less useful than B and D. So obtain knowledge about the system in water, it could prove more useful to not include coacervates. However, this could in turn cause the formation of aggregates and a comprehensive experiment of the entire system is more elegant.

Conclusion

This study presents a thermodynamic model for the partitioning of self-replicator building blocks and their oligomers. Statistical thermodynamics is a useful tool to describe partitioning of these species into complex coacervates. It allows to extract energetic parameters for different species from experimental data or to predict the behaviour based upon input for the standard chemical potential.

The simplest case of the model in which only monomer partitioning is considered gives reasonable resembles with experimental data showing that the principles of the model are founded and suggesting that the model can be used more generally. It appeared that monomers were not able to occupy many of the volume of the coacervate. An explanation has yet to be found but it is proposed that the monomer is only capable to partition into a coacervate by replacing small ions that serve for charge neutrality or are present to maximize entropy. Since the available volume of the coacervate is small, the concentration of monomer is effectively larger than when the entire volume of the coacervate is considered. This causes the partition coefficient to be in the order of 10^7.

The model is extended to include macrocycles and linear oligomers. For this, it is assumed that the standard chemical potential of cyclic and linear species follow affine functions. Then the species are distribution by an exponential relationship. This is used to learn about the prevalence of larger species when comparing the solvent and coacervate phase. For macrocycles, only small differences in the linear dependence of the standard chemical potential of 2 k_bT can cause the presence of a pentamer or hexamer relative to the total number of cycles to increase 10 to 100-fold when going from solvent to coacervate. This effect is even stronger when saturation occurs in the coacervate. For a mixture of linear and cyclic species in two phases, only the most favourable species (usually cycles in coacervate) has the possibility for the ratio of two consecutive rings to be much above 0.1, thus being able to effectively form larger species. Oxidation of dithiols causes the macrocycle and linear distribution to push to larger species but virtually this only effects rings in coacervates. Then $\tilde{B}$ can increase two-fold in going from little to almost complete oxidation.

Experimental data can be used to make fits and extract energetic parameters. This has not been done based on limited useful data for the water phase and the absence of coacervate content measurements. Careful experiments on the determination of the composition in water and coacervate can in the first place be used to validate the model. If is does, it allows to extract energetic parameters for different species allowing to gain insight into the

environment in the coacervate such as the interactions with polyelectrolytes or the effect of salt ions.

Future extensions of the model will be concerned with including diffusion behaviour and different building blocks in order to find the conditions for which it is possible to form different replicators in (separate) coacervates or in two phases. Currently it is observed that for two replicators in water, one replicator will "eat" the other over time. Coacervates might allow for spatial separation of material, or "information", such as replicators while smaller molecules can diffuse between coacervate and serve for "communication". Both spatial separation and more building blocks increases complexity of the system, allowing to include more functions such that at some point a living molecular system can be made.

References

(1) Shirt-Ediss, B.; Murillo-Sánchez, S.; Ruiz-Mirazo, K. *Beilstein journal of organic chemistry* **2017**, *13*, 1388–1395.

(2) Mann, S. *Accounts of chemical research* **2012**, *45*, 2131–2141.

(3) Ruiz-Mirazo, K.; Briones, C.; de la Escosura, A. *Chemical Reviews* **2014**, *114*, 285–366.

(4) Gánti, T., *The principles of life*; Oxford University Press: 2003.

(5) Pross, A. *Chemistry–A European Journal* **2009**, *15*, 8374–8381.

(6) Ruiz-Mirazo, K.; Peretó, J.; Moreno, A. *Origins of Life and Evolution of the Biosphere* **2004**, *34*, 323–346.

(7) Research: From self-replication towards de-novo life Accessed: 20/6/2020, `https://www.otto-lab.com/research/`.

(8) Carnall, J. M.; Waudby, C. A.; Belenguer, A. M.; Stuart, M. C.; Peyralans, J. J.-P.; Otto, S. *Science* **2010**, *327*, 1502–1506.

(9) Colomb-Delsuc, M.; Mattia, E.; Sadownik, J. W.; Otto, S. *Nature communications* **2015**, *6*, 1–7.

(10) Monreal Santiago, G. Steps towards de-novo life: compartmentalization and feedback mechanisms in synthetic self-replicating systems, Ph.D. Thesis, University of Groningen, 2020.

(11) Ottelé, J.; Hussain, A. S.; clemens mayer; Otto, S. **2019**.

(12) Sadownik, J. W.; Mattia, E.; Nowak, P.; Otto, S. *Nature chemistry* **2016**, *8*, 264.

(13) Spoelstra, W. K.; Deshpande, S.; Dekker, C. *Current opinion in biotechnology* **2018**, *51*, 47–56.

(14) Liu, X.; Haddou, M.; Grillo, I.; Mana, Z.; Chapel, J.-P.; Schatz, C. *Soft Matter* **2016**, *12*, 9030–9038.

(15) Frederix, P. W.; Idé, J.; Altay, Y.; Schaeffer, G.; Surin, M.; Beljonne, D.; Bondarenko, A. S.; Jansen, T. L.; Otto, S.; Marrink, S. J. *ACS nano* **2017**, *11*, 7858–7868.

(16) Nakashima, K. K.; Vibhute, M. A.; Spruijt, E. *Frontiers in molecular biosciences* **2019**, *6*.

(17) Dyson, F., *Origins of life*; Cambridge University Press: 1999.

(18) Poudyal, R. R.; Guth-Metzler, R. M.; Veenis, A. J.; Frankel, E. A.; Keating, C. D.; Bevilacqua, P. C. *Nature communications* **2019**, *10*, 1–13.

(19) Frankel, E. A.; Bevilacqua, P. C.; Keating, C. D. *Langmuir* **2016**, *32*, 2041–2049.

(20) Aumiller Jr, W. M.; Keating, C. D. *Nature chemistry* **2016**, *8*, 129.

(21) Mizuuchi, R.; Blokhuis, A.; Vincent, L.; Nghe, P.; Lehman, N.; Baum, D. *Chemical communications* **2019**, *55*, 2090–2093.

(22) Chandler, D., *Introduction to modern statistical mechanics*, 1987.

(23) Van Lente, J. J.; Claessens, M. M.; Lindhoud, S. *Biomacromolecules* **2019**, *20*, 3696–3703.

(24) Martin, N.; Li, M.; Mann, S. *Langmuir* **2016**, *32*, 5881–5889.

(25) McTigue, W. C. B.; Perry, S. L. *Soft Matter* **2019**, *15*, 3089–3103.

(26) McCall, P. M.; Srivastava, S.; Perry, S. L.; Kovar, D. R.; Gardel, M. L.; Tirrell, M. V. *Biophysical journal* **2018**, *114*, 1636–1645.

(27) Cruz, M. A.; Morris, D. L.; Swanson, J. P.; Kundu, M.; Mankoci, S. G.; Leeper, T. C.; Joy, A. *ACS Macro Letters* **2018**, *7*, 477–481.

(28) Schlenoff, J. B.; Yang, M.; Digby, Z. A.; Wang, Q. *Macromolecules* **2019**, *52*, 9149–9159.

(29) Hamad, F. G.; Chen, Q.; Colby, R. H. *Macromolecules* **2018**, *51*, 5547–5555.

(30) Perry, S. L.; Li, Y.; Priftis, D.; Leon, L.; Tirrell, M. *Polymers* **2014**, *6*, 1756–1772.

(31) Zhao, M.; Xia, X.; Mao, J.; Wang, C.; Dawadi, M. B.; Modarelli, D. A.; Zacharia, N. S. *Molecular Systems Design & Engineering* **2019**, *4*, 110–121.

(32) Huang, S.; Zhao, M.; Dawadi, M. B.; Cai, Y.; Lapitsky, Y.; Modarelli, D. A.; Zacharia, N. S. *Journal of colloid and interface science* **2018**, *518*, 216–224.

(33) Schaeffer, G.; Mattia, E.; Markovitch, O.; Liu, K.; Hussain, A. S.; Ottelé, J.; Sood, A.; Otto, S. **2020**, Chemical Fueling Enables Molecular Complexification of Assembly-Driven Self-Replicators.

(34) Blokhuis, A. Physical aspects of the origins of life, Ph.D. Thesis, Université Paris Sciences et Lettres, 2019.

(35) Komáromy, D.; Stuart, M. C.; Monreal Santiago, G.; Tezcan, M.; Krasnikov, V. V.; Otto, S. *Journal of the American Chemical Society* **2017**, *139*, 6234–6241.

(36) Lahiri, S.; Wang, Y.; Esposito, M.; Lacoste, D. *New Journal of Physics* **2015**, *17*, 085008.

Appendices

Pieter Brongers, *University of Groningen*

Contents

1 Models monomer

To make plots and find parameters for monomer partitioning, several methods have been used. These are outlined below. First, the formulae for plotting the observed partition constant as a function of the amount of material relative to the number of sites in the coacervate is shown. To get K_{obs} over the absolute amount of material, a different code is used. Finally, the code for finding fit parameters is shown.

1.1 General formula

The most basic description of the system is that one molecule **FK** is of similar size as a water molecule and that they can replace one another between a coacervate and the solvent.

$$\mathbf{1}\ (\text{aq}) + \text{H}_2\text{O}\ (\text{co}) \overset{\Delta\mu}{\rightleftharpoons} \mathbf{1}\ (\text{co}) + \text{H}_2\text{O}\ (\text{aq}) \tag{eq. 1}$$

For this reaction, a simple formula can be derived from statistical thermodynamics as shown in (eq. 2).

$$K = \frac{\tilde{n}(M-m)}{m(N-\tilde{n})} \approx \frac{\tilde{n}M}{m(N-\tilde{n})} = \frac{\tilde{n}M}{m(\phi M-\tilde{n})} \tag{eq. 2}$$

In here, $\tilde{n}$ and m are the number of **FK** in coacervates and water, respectively. N and M are the number of active sites in coacervates and water. $(M-m)$ is approximated as M since $M >> m$. Since the model is not elaborate at this point, this approximation is not applied yet.
This formula is rewritten and both sides are divided by n^2,

$$\tilde{n}(M-m) = Km(N-\tilde{n}) \tag{eq. 3}$$

$$\overset{/n^2}{\Rightarrow}\ \gamma_{co}\left(\frac{M}{n}-\gamma_{aq}\right) = K\gamma_{aq}\left(\frac{N}{n}-\gamma_{co}\right) \tag{eq. 4}$$

With $\gamma_x = n_x/n$ describing the fraction of **FK** that resides in medium x. The formula is rewritten in terms of n/N and M/N such that the system can be described based on the amount of **FK** added relative to the number of sites in coacervates and the amount of water sites relative to coacervates sites, respectively. This leads to (eq. 5).

$$\gamma_{co}\left(\frac{M/N}{n/N}-\gamma_{aq}\right) = K\gamma_{aq}\left(\frac{1}{n/N}-\gamma_{co}\right) \tag{eq. 5}$$

This equation can be solved in terms of γ_{co} and γ_{aq} since their sum equals 1. From this the observed partition coefficient, K_{obs}, can be determined,

$$K_{obs} = \frac{x_{co}}{x_{aq}} = \frac{\tilde{n}/N}{m/M} = \frac{\gamma_{co}\cdot n/N}{\gamma_{aq}\cdot\frac{n/N}{M/N}} \tag{eq. 6}$$

The set of equations (eq. 5) and $\gamma_{co}+\gamma_{aq}=1$ are solved using the *SymPy* option `solve` (imported as `solve1`). The corresponding code is shown in figure 1. This is used to make $K_{obs}, n/N$-plots.

1.2 Comparison with experiment

In order to calculate K_{obs} as a function of n (eq. 2) is rewritten. Since $K_{obs} = x_{co}/x_{aq}$, expressing (eq. 2) in terms of x_{co}, while using that $x_{aq} = n/m - \phi x_{co}$, results in (eq. 7) that is solved using `sympy.solvers.solve`.

$$n^*(M-m) = Km(N-n^*) \Rightarrow x_{co}(\frac{1}{\phi}-\frac{n}{\phi M}+x_{co}) = K(\frac{n}{\phi M}-x_{co})(1-x_{co}) \tag{eq. 7}$$

```
def l(nN, MN, K):
    sol = solve1([GAMco*((MN / nN)-GAMaq) - K*GAMaq*(1 / nN - GAMco), GAMco + GAMaq - 1], GAMco, GAMaq)
    for i in range(len(sol)):
        if float(sol[i][0]) > 0:
            if float(sol[i][1]) > 0:
                break
    fracn_inco = sol[i][0]
    fracn_inaq = sol[i][1]
    xco = fracn_inco * nN
    xaq = fracn_inaq * nN / MN
    return [fracn_inco, fracn_inaq, xco, xaq, xco/xaq]

def K(neg_nN, pos_nN, MN, K, nN_increment):
    Kapp = []
    range_nN = logspace(-1*neg_nN, pos_nN, nN_increment)
    for i_nN in range_nN:
        results = l(i_nN, MN, K)
        Kapp.append(results[4])
    return [list(range_nN),Kapp]
```

Figure 1. Code that is used to calculate the observed partition constant in terms of relative amount of material and sites.

```
def k(n, phi, K):
    sol = solve1(Xco * (1/phi - n/(phi*M) + Xco) - K * (n/(phi*M) - Xco) * (1-Xco), Xco)
    for i in range(len(sol)):
        if sol[i] > 0:
            if sol[i] < 1:
                break
    xco = sol[i]
    xaq = n/M - phi * sol[i]
    return [xco, xaq, xco/xaq]

def graph(neg_n, pos_n, phi, K, nincrement):
    Kapp = []
    range_n = logspace(-1*neg_n, pos_n, nincrement)
    for i_n in range_n:
        results1 = k(i_n, phi, K)
        Kapp.append(results1[2])
    return [list(range_n),Kapp]
```

Figure 2. Code for calculating K_{obs} as a function of n, K and ϕ.

Finding fit parameters

In order to find fit parameters for K, ϕ and ζ, the expression for K is rewritten into a function for m in (eq. 8) and (eq. 9). Fit parameters are found using the *SciPy* option `curve_fit` (imported as `nlfit`) (fig. 3,4). An initial guess is provided in *p0*.

$$K = \frac{\tilde{n}(M-m)}{m(\phi M - \tilde{n})} \approx \frac{\tilde{n}M}{m(\phi M - \tilde{n})} \Rightarrow m = \frac{\tilde{n}M}{K(\phi M - \tilde{n})} \qquad \text{(eq. 8)}$$

$$K = \frac{\tilde{n}(1-(\zeta-1)\frac{\tilde{n}}{\phi M})^{\zeta-1}}{m\phi(1-\zeta\frac{\tilde{n}}{\phi M})^{\zeta}} \Rightarrow m = \frac{\tilde{n}(1-(\zeta-1)\frac{\tilde{n}}{\phi M})^{\zeta-1}}{K\phi(1-\zeta\frac{\tilde{n}}{\phi M})^{\zeta}} \qquad \text{(eq. 9)}$$

```
def k(n_star, phi, K):
    return (n_star * M)/((phi*M - n_star)*K)

popt, pcov = nlfit(k, nstar, m, p0=(0.015, 1000))
```

Figure 3. Code for finding fit parameters when a monomer is regarded the same size as a water molecule.

```
def k(n_star, zeta, phi, K):
    return (n_star * (1 - (zeta - 1)*n_star/(phi*M))**(zeta-1))/(K * phi * (1 - zeta*n_star/(phi*M))**zeta)

popt, pcov = nlfit(k, nstar, m, p0=(40, 0.015, 1000))
```

Figure 4. Code for obtaining fit parameters while taking into account the size of a monomer compared to H_2O.

2 Experimental data

Monomer partitioning has been monitored at different concentrations for various times after addition. The resulting capacity factors are shown below in a table and plot. The green line in the plots relates to the highest value of monomer in coacervate ($\tilde{n}$) from the data points.

Monomer conc. (mM)	**0.05**	**0.2**	**0.5**	**1**	**2**
10 min.	257.200	385.377	3.85373	1.92478	0.760010
60 min.	123.627	67.0081	5.90268	2.87115	1.77135
180 min.	265.973	32.7331	7.19788	7.23095	1.62095
1 day	–	–	12.3715	4.57195	1.99366
2 days	577.557	867.460	30.7231	2.47558	0.100741

Table 1. Capacity factor for different concentrations of monomers, measured at various time scales after addition of monomer.

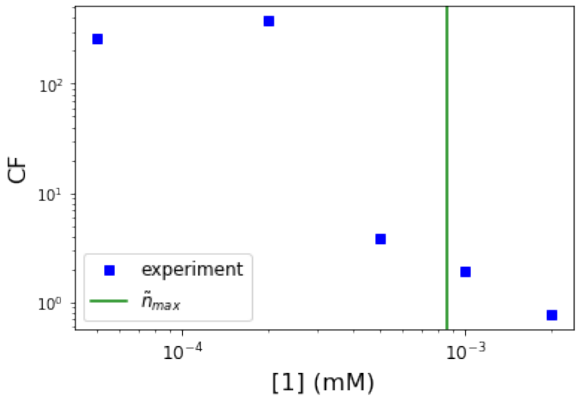

Figure 5. Capacity factor measured 10 minutes after addition of monomers.

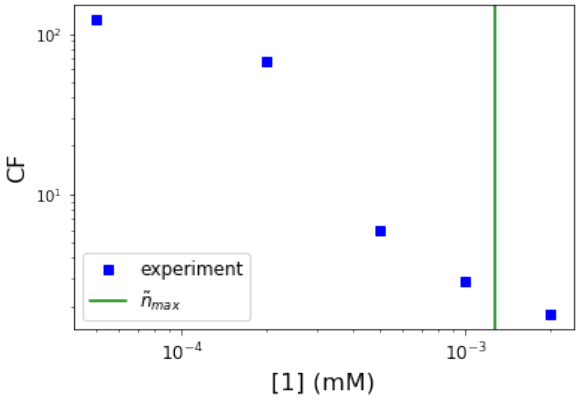

Figure 6. Capacity factor measured 60 minutes after addition of monomers.

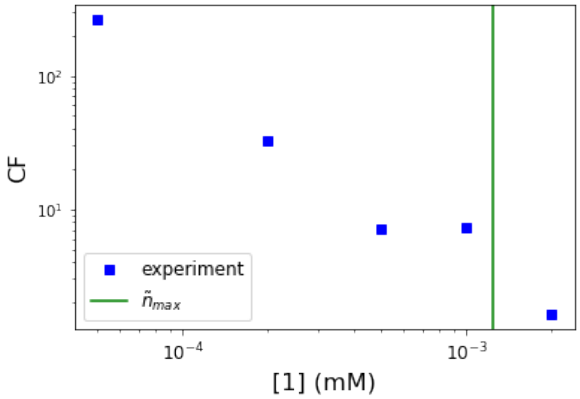

Figure 7. Capacity factor measured 180 minutes after addition of monomers.

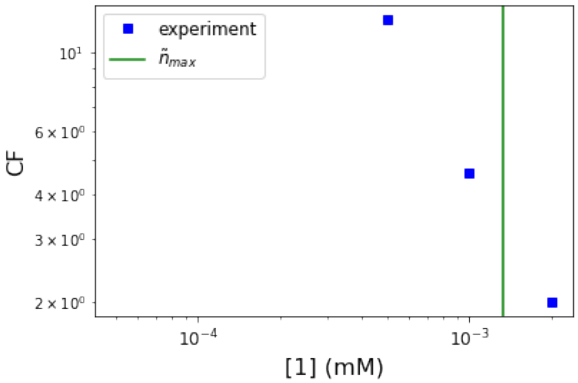

Figure 8. Capacity factor measured 1 day after addition of monomers.

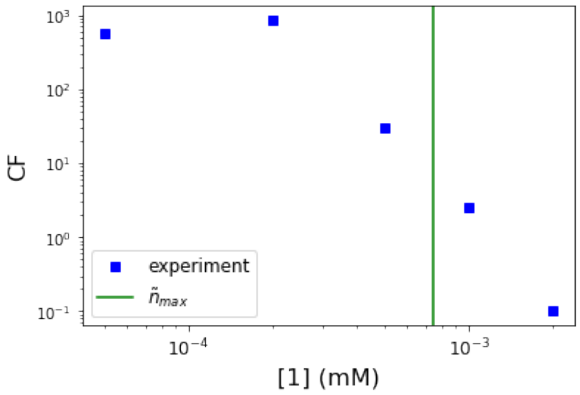

Figure 9. Capacity factor measured 2 days after addition of monomers.

3 Macrocycle partitioning

3.1 Derivations

This section contains some of the derivations that have been mentioned rather than worked out in the main article. The relationships are useful and might not be visible at first sight and for this reason they are derived here.

3.1.1 Exponential relationship macrocycles

An exponential function can be used to describe the distribution of interconverting species, in this case macrocycles and later on, linear species as well. The parameters cannot always be extracted in terms of constants or other known values. However it is possible in the case of a species in a dilution limit. In that case, the partition constant does not contain factors other that the concentrations of species participating in the reaction. The following derivation regards macrocycles in water.

$$[\mathrm{m}] + [\mathrm{n}] \overset{\mathrm{K}}{\rightleftharpoons} [\mathrm{m{+}n}]$$

$$\text{for k=3, } \chi_3 = AB^{3-3} = A$$

$$K = \frac{\chi_{m+n}}{\chi_m \chi_n}$$

$$K \cdot AB^{m-3} \cdot AB^{n-3} = AB^{m+n-3}$$

$$K \cdot AB^{-3} = 1$$

$$B = (AK)^{1/3} = (\chi_3 K)^{1/3}$$

$$\text{in conclusion: } \chi_k = \chi_3 \left((\chi_3 K)^{1/3}\right)^{k-3}$$

3.1.2 Species distribution in water and coacervate

Similarly to the previous derivation, the dilution limit allows for more analytic expressions. Then, the partition constant is the ratio of concentrations in two phases and the exponential parameters in one phase can be expressed in terms of the ones in the other phase when including energetic differences. In here, it is shown for cyclic species in water and coacervate.

Cyclic species	Linear species
$\chi_k = AB^{k-3}$	$\chi_k^* = CD^{k-1}$
$K_k = \frac{\theta_k}{\chi_k} \Rightarrow \theta_k = \chi_k K_k$	$K_k^* = \frac{\theta_k^*}{\chi_k^*} \Rightarrow \theta_k^* = \chi_k^* K_k^*$
$\theta_k = AB^{k-3} e^{-\beta((\tilde{a}-a)+(\tilde{b}-b)k)}$	$\theta_k^* = CD^{k-1} e^{-\beta((\tilde{a}^*-a^*)+(\tilde{b}^*-b^*)k)}$
$\theta_k = AB^{k-3} e^{-\beta((\tilde{a}-a)+(\tilde{b}-b)3)} e^{-\beta(\tilde{b}-b)(k-3)}$	$\theta_k^* = CD^{k-1} e^{-\beta((\tilde{a}^*-a^*)+(\tilde{b}^*-b^*)1)} e^{-\beta(\tilde{b}^*-b^*)(k-1)}$
$\theta_k = AK_3 \cdot B^{k-3} e^{-\beta(\tilde{b}-b)(k-3)} = \tilde{A}\tilde{B}^{k-3}$	$\theta_k^* = CK_1^* \cdot D^{k-1} e^{-\beta(\tilde{b}^*-b^*)(k-1)} = \tilde{C}\tilde{D}^{k-1}$
$\tilde{A} = AK_3$ and $\tilde{B} = Be^{-\beta(\tilde{b}-b)}$	$\tilde{C} = CK_1^*$ and $\tilde{D} = De^{-\beta(\tilde{b}^*-b^*)}$

Table 2. Using the equation for partitioning, the distribution of a species in one phase can be expressed in terms of the distribution in the other phase. This is only applicable in the case of a diluted phase.

4 Models macrocycles

4.1 Introduction

In the case of larger systems, i.e. more species and/or saturation, it is not always possible to derive expressions analytically and use those in a notebook environment. For this reason, the parameters in the exponential distributions of species can be expressed in terms of one of them and the conversion factors are determined by the energetic parameters of the systems - which embody the input of the computation. This allows one to solve for one parameter, infer the others and thus know all concentrations and ratios.
All species are considered to be distributed exponentially over length, both in the coacervate and water. Parameters for cycles are A and B and for linear species C and D, a tilde indicating the coacervate phase. Thus,

$$\begin{aligned} \text{cyclic fraction in water} \quad & \chi_k = AB^{k-3} \\ \text{linear fraction in water} \quad & \chi_k^* = CD^{k-1} \\ \text{cyclic fraction in coacervate} \quad & \theta_k = \tilde{A}\tilde{B}^{k-3} \\ \text{linear fraction in coacervate} \quad & \theta_k^* = \tilde{C}\tilde{D}^{k-1} \end{aligned}$$

In virtually all the models that have been constructed, the parameters are expressed in terms of B, for the sake of simplicity and clarity, as the relative distribution of macrocycles in a diluted water phase is a natural reference frame.
In the upcoming sections, for different systems, the expressions for the various parameters in terms of B are given. They have been applied in this form in the software.

4.2 Linear and cyclic species; Oxidation

Both linear and cyclic species have been considered. The linear species can interconvert. They can also react with macrocycles, forming one new linear molecule. This can be considered as a reaction step in the mechanism of ring conversion with catalytic contribution of the linear species. Relevant exchange reactions are:

$$\begin{aligned} [\mathrm{m}]^* + [\mathrm{n}]^* &\overset{\mathrm{K_{ex,l}}}{\rightleftharpoons} [\mathrm{m\text{-}k}]^* + [\mathrm{n+k}]^* \\ [\mathrm{m}] + [\mathrm{n}]^* &\overset{\mathrm{K_{ex,m}}}{\rightleftharpoons} [\mathrm{m+n}]^* \end{aligned}$$

In the coacervate these are:

$$\begin{aligned} [\tilde{m}]^* + [\tilde{n}]^* &\overset{\tilde{K}_{\mathrm{ex,l}}}{\rightleftharpoons} [\tilde{m}-\tilde{k}]^* + [\tilde{n}+\tilde{k}]^* \\ [\tilde{m}] + [\tilde{n}]^* &\overset{\tilde{K}_{\mathrm{ex,m}}}{\rightleftharpoons} [\tilde{m}+\tilde{n}]^* \end{aligned}$$

The equilibrium coefficient for exchange of linear species is 1, and for the cyclic+linear exchange it is dependent on the size of the macrocycles such that $K_{ex,m} = EF^m$. This is shown in table 3.
The partition reactions that need to be considered are:

$$\begin{aligned} [\mathrm{m}] &\overset{\mathrm{K}_m}{\rightleftharpoons} [\tilde{m}] \\ [\mathrm{m}]^* &\overset{\mathrm{K}_m^*}{\rightleftharpoons} [\tilde{m}]^* \end{aligned}$$

Exchange linear species	Exchange linear and cyclic species
$[\mathrm{m}]^* + [\mathrm{n}]^* \overset{\mathrm{K_{ex,l}}}{\rightleftharpoons} [\mathrm{m\text{-}k}]^* + [\mathrm{n+k}]^*$	$[\mathrm{m}] + [\mathrm{n}]^* \overset{\mathrm{K_{ex}}}{\rightleftharpoons} [\mathrm{m+n}]^*$
$K_{ex,l} = exp[-\beta\Delta\mu^\circ]$	$K_{ex,m} = exp[-\beta\Delta\mu^\circ]$
$\Delta\mu^\circ = [\mu^{\circ *}_{m-k} + \mu^{\circ *}_{n+k}] - [\mu^{\circ *}_m + \mu^{\circ *}_n]$	$\Delta\mu^\circ = [\mu^{\circ *}_{m+n}] - [\mu^\circ_m + \mu^{\circ *}_n]$
$\Delta\mu^\circ = [a^* + b^*(m-k) + a^* + b^*(n+k)]$	$\Delta\mu^\circ = [a^* + b^*(m+n)] - [a + bm + a^* + b^* n]$
$-[a^* + b^* m + a^* + b^* n] = 0$	$\Delta\mu^\circ = b^* m - a - bm$
hence $K_{ex,l} = 1$	hence $K_{ex,m} = exp[-\beta(-a + b^* m - bm)]$
	$K_{ex,m} = exp[\beta a] \cdot exp[-\beta(b^* - b)m] = EF^m$

Table 3. Derivation for the equilibrium constant for two different exchange reactions.

Conservation laws

Two conservation laws can be considered:

- Total mass

$$\begin{aligned} W &= M\sum_{l=1}^{\infty} l\cdot\chi_l^* + M\sum_{l=3}^{\infty} l\cdot\chi_l + N\sum_{l=1}^{\infty} l\cdot\theta_l^* + N\sum_{l=3}^{\infty} l\cdot\theta_l \\ &= M\sum_{l=1}^{\infty} l\cdot CD^{l-1} + M\sum_{l=3}^{\infty} l\cdot AB^{l-3} + N\sum_{l=1}^{\infty} l\cdot\tilde{C}\tilde{D}^{l-1} + N\sum_{l=3}^{\infty} l\cdot\tilde{A}\tilde{B}^{l-3} \\ \frac{W}{M} &= \frac{C}{(D-1)^2} + \frac{A(3-2B)}{(B-1)^2} + \frac{\tilde{C}\phi}{(\tilde{D}-1)^2} + \frac{\tilde{A}\phi(3-2\tilde{B})}{(\tilde{B}-1)^2} \end{aligned}$$

- Total number of linear species

$$\begin{aligned} Z &= M\sum_{l=1}^{\infty}\chi_l^* + N\sum_{l=1}^{\infty}\theta_l^* \\ &= M\sum_{l=1}^{\infty} CD^{l-1} + N\sum_{l=1}^{\infty}\tilde{C}\tilde{D}^{l-1} \\ \frac{Z}{M} &= \frac{C}{1-D} + \frac{\tilde{C}\phi}{1-\tilde{D}} \end{aligned}$$

From the second conservation law,

$$C = \frac{Z}{M}\left(\frac{1}{1-D} + \frac{K_1^*\phi}{1-\tilde{D}}\right)^{-1}$$

Oxidation level

The extent of oxidation can be defined in the following way,

$$oxi = \frac{\#\text{S in RSSR}}{\#\text{S in RSSR} + \#\text{S in RSH}} = \frac{2W - 2Z}{2W} = 1 - \frac{Z}{W}$$

and thus,

$$Z = (1 - oxi)W$$

Parameters

All parameters can be expressed in terms of B using the derivation described above and in section 3.1.

$$A = \frac{1}{E}B^3 \qquad \tilde{A} = A \cdot K_3$$

$$B = B \qquad \tilde{B} = B \cdot e^{-\beta(\tilde{b}-b)}$$

$$C = \frac{(1-oxi)W}{M}\left(\frac{1}{1-D} + \frac{K_1^*\phi}{1-\tilde{D}}\right)^{-1} \qquad \tilde{C} = C \cdot K_1^*$$

$$D = B \cdot F \qquad \tilde{D} = D \cdot e^{-\beta(\tilde{b}^*-b^*)}$$

In here,

$$E = e^{\beta a}$$

$$F = e^{-\beta(b^*-b)}$$

$$K_3 = e^{-\beta((\tilde{a}-a)+(\tilde{b}-b)3)}$$

$$K_1^* = e^{-\beta((\tilde{a}^*-a^*)+(\tilde{b}^*-b^*)1)}$$

β is incorporated in the standard chemical potentials that are expressed in terms of k_bT.

Code

The previously mentioned parameters are incorporated in a code in which B will be determined and the level of oxidation can be varied. Values for the parameters in the standard chemical potential function have to be specified as well. The code is shown in figure 10. In here, B is found by using the conservation law for the total mass. This only contains preset constants or B-dependent parameters. Getting all terms to the same side of the equality means that the entire expression equals zero. A numerical approach is applied in which B is changed until it converges. From the result, $\tilde{B}$, D and $\tilde{D}$ are also determined and this can be done for the other parameters as well.

```
def g(oxi): # calculate B for a level of oxidation
    def gg(B):
        A = (1/E)*B**3
        Ac = A * K3
        Bc = B * exp(-1*(bcc-bcw))
        D = B * F
        Dc = D * exp(-1*(blc-blw))
        C = (W/M)*(1-oxi)*(1/((1-D)**(-1)+Ks1*phi*(1-Dc)**(-1)))
        Cc = C * Ks1
        return C/((D-1)**2) + (A*(3-2*B))/((B-1)**2) + Cc*phi/((Dc-1)**2) + (Ac*phi*(3-2*Bc))/((Bc-1)**2) - W/M

    tolerance = 1e-6
    x0,x1 = 0,0.999
    xl,xr = x0,x1
    sl,sr = sign(gg(xl)),sign(gg(xr))
    while (xr - xl) > tolerance:
        xm = (xl + xr) / 2.
        sm = sign(gg(xm))
        if sm == sr:
            xr,sr = xm,sm
        else:
            xl,sl = xm,sm
    Bcfinal = xm * exp(-1*(bcc-bcw))
    Dfinal = xm * F
    Dcfinal = Dfinal * exp(-1*(blc-blw))
    return [xm, Bcfinal, Dfinal, Dcfinal]
```

Figure 10. The exponential parameters can be determined as a function of the level of oxidation if the energetic variables are given.

```
W = 0.001
phi = 0.00001
M = 1.39
acw = 2       # cycle in water
acc = acw-1   # cycle in coacervate
alw = 2       # linear in water
alc = alw-1   # linear in coacervate
bcw = -2.5    # cycle in water
bcc = bcw-2   # cycle in coacervate
blw = -1      # linear in water
blc = blw-2   # linear in coacervate
E = exp(acw)
F = exp(-1*(blw-bcw))
K3 = exp(-1*(acc-acw+(bcc-bcw)*3))
Ks1 = exp(-1*(alc-alw+(blc-blw)*1))

plots(W)
```

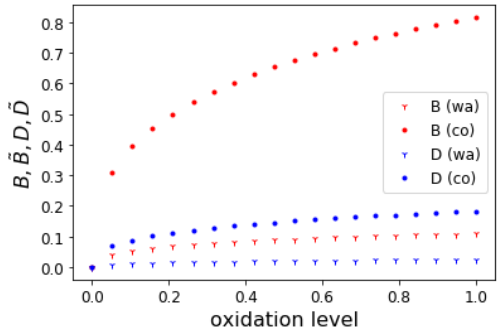

Figure 11. Input parameters and plot as shown in the report.

4.3 System at saturation

In order to analyse a system at saturation, the methods applied before do not work entirely. That is, it is not possible to express the exponential parameters in both water and coacervate in terms of B. The issue originates from the equilibrium condition for partitioning. In stead of solely the concentration of the species that partitions, it also contains θ_W and θ_0 such that,

$$K_k = \frac{\theta_k}{\chi_k} \left(\frac{\theta_W}{\theta_0} \right)^{k-1} \frac{1}{\theta_0}$$

This leads to parameters that are expressed in terms of one another in a loop. That is why in this case only one saturated phase will be considered. Then, everything can be expressed in terms of $\tilde{B}$.

4.3.1 Cyclic species

The equilibrium constant for the exchange reaction contains an additional term θ_W. This removes the possibility to liberate B and find the function for θ_k as in section 3.1.1. Though, one can express A in terms of B by considering that K is constant and given by:

$$\tilde{K} = \frac{\theta_{m+n}}{\theta_m \theta_n} \theta_W$$

and

$$\theta_W = 1 - \sum_{i=3}^{\infty} (i-1)\theta_i = 1 - \sum_{i=3}^{\infty} (i-1)\tilde{A}\tilde{B}^{i-3} = 1 - \frac{\tilde{A}(2-\tilde{B})}{(\tilde{B}-1)^2}$$

Then,

$$\tilde{K} = \frac{\tilde{B}^3}{\tilde{A}} \cdot \theta_W = \frac{\tilde{B}^3}{\tilde{A}} - \frac{\tilde{B}^3(2-\tilde{B})}{(\tilde{B}-1)^2}$$

$$\Rightarrow \tilde{A} = \left(\frac{K_{ex}}{\tilde{B}^3} + \frac{2-\tilde{B}}{(\tilde{B}-1)^2} \right)^{-1}$$

Another relevant quantity is θ_0. Its derivation is similar to θ_W,

$$\theta_0 = 1 - \sum_{i=3}^{\infty} i \cdot \theta_i = 1 - \left(\frac{\tilde{A}(3-2\tilde{B})}{(\tilde{B}-1)^2} \right)$$

These expressions are implemented in a notebook. The maximum of the graph of A is found by a numerical approach. The procedure initiates with two values for B, $B_{left} = 0.5$ and $B_{right} = 0.999$. For the average, B, A is calculated, as well as for $B + \delta$, with small delta. For $A(B) < A(B+\delta)$, the average B becomes the new B_{left}, i.e. $B \rightarrow B_{left}$ and for $A(B) > A(B+\delta)$, $B \rightarrow B_{right}$. A new average is taken and the process is repeated until a tolerance level of $B_{right} - B_{left} < 1E(-6)$ is reached.
The final code is shown in figure 12. In there, also θ_0 can be calculated.

4.3.2 Cyclic and linear species

The procedure for a system of linear and cyclic species is similar to the one in the previous section. It differs by the addition of a sum over the linear material such that,

$$\theta_W = 1 - \sum_{i=1}^{\infty} (i-1)\theta_i^* - \sum_{i=3}^{\infty} (i-1)\theta_i$$

$$\theta_0 = 1 - \sum_{i=1}^{\infty} i \cdot \theta_i^* - \sum_{i=3}^{\infty} i \cdot \theta_i$$

This results in,

$$\tilde{A} = \left(1 - \frac{\tilde{C}\tilde{D}}{(\tilde{D}-1)^2}\right)\left(\frac{\tilde{E}}{\tilde{B}^3} + \frac{2-\tilde{B}}{(\tilde{B}-1)^2}\right)^{-1}$$
$$\theta_0 = 1 - \left(\frac{\tilde{A}(3-2\tilde{B})}{(\tilde{B}-1)^2}\right) - \left(\frac{\tilde{C}}{(\tilde{D}-1)^2}\right)$$

with,

$$\tilde{C} = \frac{(1-oxi)W}{M}(1-\tilde{D})$$
$$\tilde{D} = \tilde{B}\cdot\tilde{F}$$
$$\tilde{F} = e^{-\beta(\tilde{b}^*-\tilde{b})}$$

Since the case is fairly similar to saturation of only cyclic species the code in figure 13 relies on the same principles as in figure 12.

```
def Aparam(B, E): # A as a function of B while considering saturation
    return 1/(E/(B**3) + (2-B)/((B-1)**2))

def Aparamdil(B, E): # A as a function of B in dilution limit
    return B**3/E

def Amax(E): # determine the maximum in Aparam(B,E)
    tolerance = 1e-6
    deltax = 1e-7
    x0,x1 = 0.5,0.999
    xl,xr = x0,x1
    sl,sr = Aparam(xl,E),Aparam(xr,E)
    while (xr - xl) > tolerance:
        xm = (xl + xr) / 2.
        xmh = xm + deltax
        sm = Aparam(xm,E)
        smh = Aparam(xmh,E)
        if sm > smh:
            xr,sr = xm,sm
        else:
            xl,sl = xm,sm
    Amaxfinal = Aparam(xm,E)
    return [xm, Amaxfinal]

def theta0(B, E): # remaining solvent content
    A = Amax(E)[1]
    return 1 - (A*(3-2*B))/((B-1)**2)
```

Figure 12. Code for plotting $\tilde{A}$ as a function of $\tilde{B}$ in the case of only cyclic species.

```python
def Aparam2(B, E, F, oxi): # A as a function of B while considering saturation and two different species
    D = B * F
    C = (1-oxi)*(1-D)*(W/M)
    A = (1 - (C*D)/((D-1)**2)) * (1/(E/(B**3) + (2-B)/((B-1)**2)))
    return A

def Amax2(E, F, oxi): # determine the maximum in Aparam2(B,E,F,oxi)
    tolerance = 1e-6
    deltax = 1e-7
    x0,x1 = 0.1,0.999
    xl,xr = x0,x1
    sl,sr = Aparam2(xl,E,F,oxi),Aparam2(xr,E,F,oxi)
    while (xr - xl) > tolerance:
        xm = (xl + xr) / 2.
        xmh = xm + deltax
        sm = Aparam2(xm,E,F,oxi)
        smh = Aparam2(xmh,E,F,oxi)
        if sm > smh:
            xr,sr = xm,sm
        else:
            xl,sl = xm,sm
    Amaxfinal = Aparam2(xm,E,F,oxi)
    return [xm, Amaxfinal]

def theta02(B, E, F, oxi): # calculate solvent content
    D = B * F
    C = (1-oxi)*WM*(1-D)
    A = Aparam2(B, E, F, oxi)
    return 1 - (A*(3-2*B))/((B-1)**2) - C/((D-1)**2)
```

Figure 13. Code for plotting $\tilde{A}$ as a function of $\tilde{B}$ when linear and cyclic species and saturation are considered.

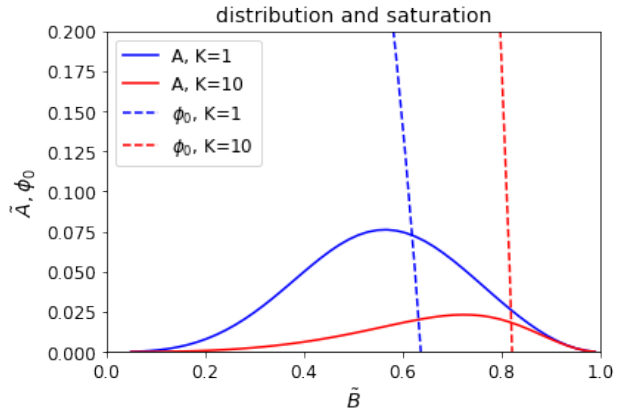

Figure 14. $\tilde{A}$ and θ_0 as function of $\tilde{B}$. The maximum for $\tilde{A}$ does not coincide with the point at which $\theta_0 = 0$

```
#parameters
F = exp(-1*1.5)
E = exp(1)
M = 1.39
W = 0.001

oxidata = linspace(0,1,20)
Blist = []
Dlist = []
Blistdil = []
Dlistdil = []
for i in oxidata:
    Blist.append(g(i)[0])
    Dlist.append(g(i)[1])
    Blistdil.append(gd(i)[0])
    Dlistdil.append(gd(i)[1])

plot(oxidata, Blist, 'r.', label='B')
plot(oxidata, Dlist, 'b.', label='D')
plot(oxidata, Blistdil, 'r1', label='B (dil)')
plot(oxidata, Dlistdil, 'b2', label='D (dil)')
xlabel('oxidation level', size='16')
ylabel('$B,D$', size='16')
xticks(size='12')
yticks(size='12')
legend(fontsize='12')
show()
```

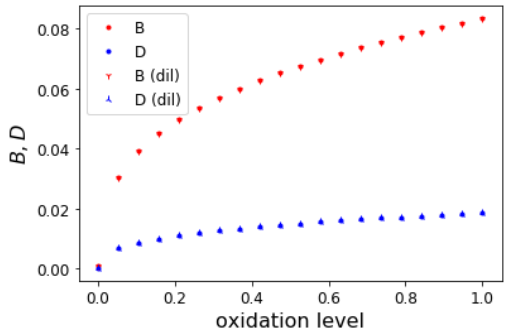

Figure 15. Distribution parameters for cyclic and linear species in water in the absence of a coacervate phase.