



Cite this: DOI: 10.1039/d5sc04365d

 All publication charges for this article have been paid for by the Royal Society of Chemistry

# Decoding information entropy of fatty acid and phospholipid vesicles *via* ordering combinatorial output of hydrazones

Reena Yadav,<sup>†</sup> Niranjani Adikessavane,<sup>†</sup> Rishi Ram Mahato and Subhabrata Maiti \*

Leveraging information entropy to quantitatively measure the organizational diversity and complexity of different chemical systems is a compelling need for next-generation supramolecular and systems chemistry. It can also be a strategy for digitalizing and enabling the bottom-up development of life-like complex systems following probable origin-of-life scenarios. According to the lipid world hypothesis, lipid molecules appear first to facilitate compartmentalization, catalysis, information processing, etc. It is envisaged that fatty acid-based vesicles are more primitive than phospholipid vesicles. Herein, we decode the difference in information storage capability of a fatty acid (oleic acid, (OA)) and a phospholipid (1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC)) vesicle by measuring vesicle-templated formation of nine different hydrazones through permutations and hierarchical ordering of combinatorial matrices involving three aldehydes and three hydrazines by determining Shannon entropy and the Gini coefficient at the systems level. This signifies a higher diversity and lower selectivity towards successful chemical reactions in OA vesicles, whereas DOPC vesicles are more selective and less diverse. Exploiting information theory in combinatorial supramolecular synthesis and unraveling information capacity relevant to cell membrane evolution will be important in understanding the information dynamicity of different transient and self-propagated synthetic and natural assembly processes over time.

Received 14th June 2025  
Accepted 21st August 2025

DOI: 10.1039/d5sc04365d

rsc.li/chemical-science

## Introduction

Information theory-derived analysis of complex phenomena has wide applications in diverse branches of *in vitro*, *in vivo* and *in silico* studies.<sup>1–4</sup> Since the genesis of Shannon's concept of information entropy in 1948, life's origin, evolution and diverse functional aspects can be interpreted by quantifying the flow of information and degree of order in biological systems.<sup>5–9</sup> Through information theory, researchers are trying to address the transition from a disordered, high-entropy state of free organic molecules to a more ordered, low-entropy state of complex biological systems for sustaining life. In other words, it is important to recognize how information was encoded in prebiotic systems and subsequently transformed with biological evolution involving a hierarchical progression to increasingly complex forms of life.

So far, utilization of information entropy from the perspective of experimental chemistry is only restricted to the analysis of molecular structures, arrays, distribution of nanoparticles *etc.*<sup>10–19</sup> To date, exploitation of information theory towards interpreting the emergence of complex systems or the

difference in information storage capacity between simple and complex systems (relevant to the evolution of biological structures) has not been experimentally delineated to the best of our knowledge. However, in the context of advancing the field of supramolecular and systems chemistry in conjunction with understanding the benefit of complexification of organic matter, the merging of information theory with adaptive chemistry of different self-organized structures is a dire necessity.<sup>20–23</sup>

J. M. Lehn exemplified Landauer's concept of 'information is physical' by defining higher molecular recognition at a constant temperature as a low entropy event, using the equation:  $I \log W = NT$  (where  $I$  = information,  $W$  = number of states,  $N$  = number of particles, and  $T$  = temperature).<sup>20,24</sup> However, this concept remained restricted to generating information patterns of molecular binding events. It has not been used to quantify the information *via* measuring Shannon entropy or the Gini coefficient through combinatorial chemistry.<sup>20,21,25</sup> It is worth mentioning that the Gini coefficient and Shannon entropy are measures of inequality or impurity and randomness or uncertainty in a probability distribution from a large dataset.<sup>5,26</sup> Chemical biologists recently employed Gini coefficients for gene-profiling, quantifying selectivity of chemical probes and small molecules towards clinically relevant target RNA or proteins.<sup>27–29</sup> In principle, the application of these statistical tools in a dataset of microcompartmental environment-specific

Department of Chemical Sciences, Indian Institute of Science Education and Research (IISER) Mohali, Knowledge City, Manauli-140306, India. E-mail: smaiti@iisermohali.ac.in

<sup>†</sup> Both these authors contributed equally.



amplification or decrease in chemical reactivity from multiple combinatorial networks can interpret the chemical information storage capacity of a given microenvironment.

Herein, we aim to explore the information entropy of two different vesicular systems, namely, fatty acid (using oleic acid (OA)) and phospholipid vesicles (using 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC)). We used vesicle-templated variability in hydrazone formation with a hierarchically ordered combinatorial matrix using three different aldehydes and hydrazines

(Fig. 1). We would like to remark that both fatty acid and phospholipid-based vesicles are attractive model protocell candidates and used as reactors for diverse biochemical processes, such as – non-enzymatic RNA synthesis, different enzymatic processes, self-sorting of supramolecular assemblies, *etc.*<sup>30–34</sup> Notably, the Lipid World hypothesis proposes that lipids and other amphiphiles played a crucial role in the origin of life by forming compartments, facilitating catalytic reactions, thereby enabling information processing.<sup>35–42</sup> This scenario

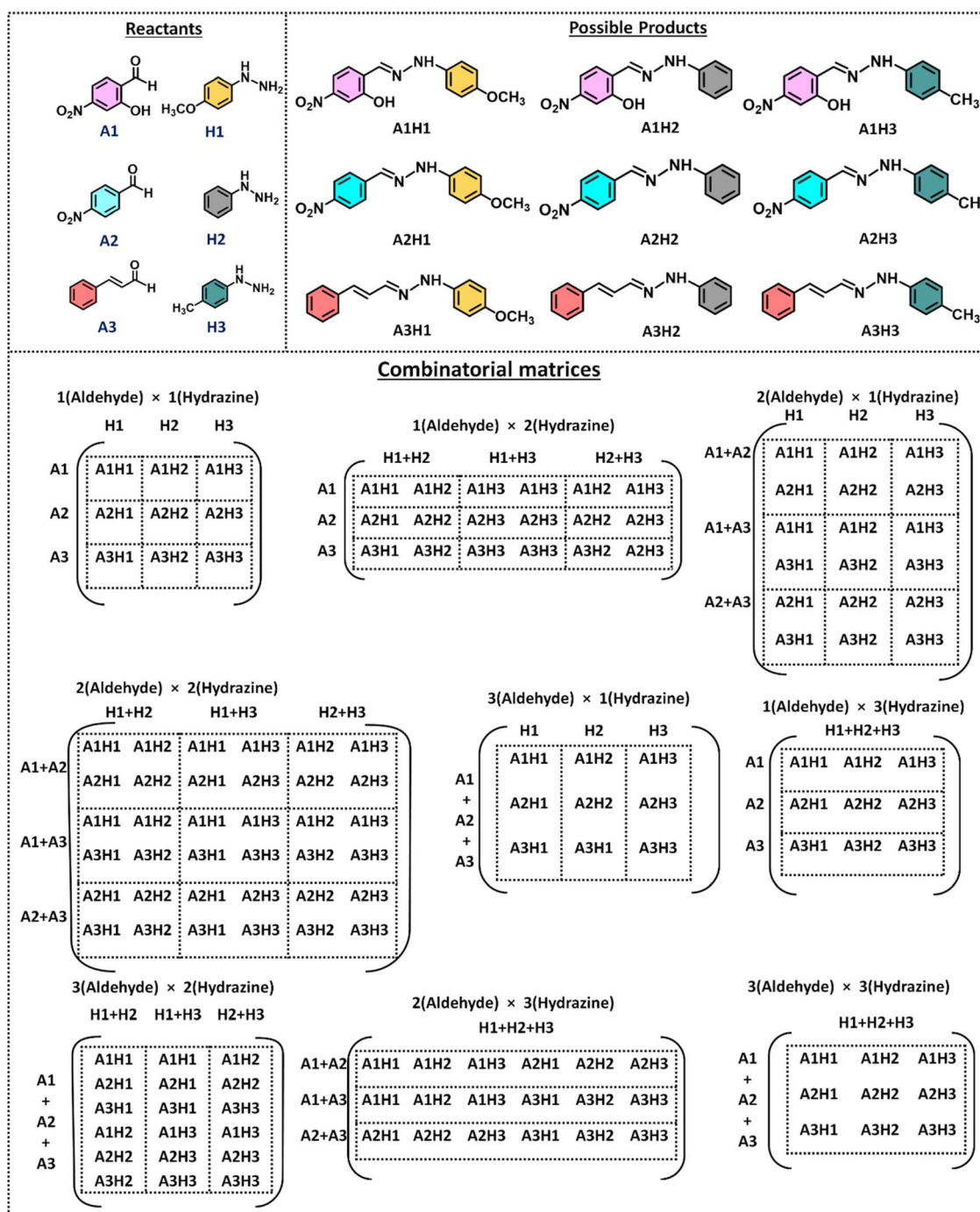


Fig. 1 Structures of 3 aldehydes and 3 hydrazine reactants and all possible 9 hydrazone products. The lower panel demonstrates all possible combinatorial matrices of different orders used in this study.



suggests that, before the emergence of complex biomolecules such as DNA and proteins, lipids and other similar molecules could have self-organized into structures such as micelles and bilayers, providing a framework for early life processes. It is also hypothesized that fatty acids, being simple, were likely present on early Earth, while phospholipids, more complex and capable of forming stable bilayers, represent a later evolution.<sup>30,35–38</sup> To this end, chemists are inclined to follow chemical reactivity in vesicular microcompartments, as it allows control of reaction conditions by mimicking the compartmentalization of biological cells, enabling simultaneous synthetic or biochemical reactions (enzymatic reaction to DNA replication) with shared intermediates for developing artificial cell-like systems.<sup>30–32,40</sup> DOPC vesicle-templated combinatorial chemistry has been used for the selective partitioning of different library members between the vesicle's lipid bilayer and the surrounding aqueous solution, leading to amplification or diminishing of certain products or selective signal transduction through the membrane from a mixture.<sup>41,42</sup> Moreover, different synthetic surfactants have been used for designing nucleotide-templated vesicular systems for temporal control over chemical reactivity, including selective hydrazone formation from a mixture.<sup>43–47</sup>

It has not yet been investigated how distinct vesicles made of various lipids might influence the distribution pattern of a product, especially when they come from a variety of combinatorial matrices arranged hierarchically. Furthermore, no research has used the outcome of combinatorial reactions to define a specific vesicle's information capacity. The two aspects described above motivated us to design a comparative combinatorial study between two different vesicles (OA and DOPC) to estimate their stored information entropy. It is to be noted here that OA and DOPC are single- and double-chain fatty acids, respectively, and have distinctly different properties in terms of their stability, permeability, and rigidity or liquid crystallinity of the hydrophobic leaflet.<sup>30,31,48</sup> Therefore, knowledge of the information capacity of these two physicochemically distinct naturally occurring vesicles will also be useful in designing synthetic systems of desired information capacity. All these facts inspired us to develop a study *via* the combinatorial outcome of hydrazone (product) distribution through differently ordered matrix inputs (reagents) to find out the information entropy (Shannon entropy and Gini coefficient) of OA and DOPC vesicles on an individual basis.

## Results and discussion

Vesicles have been utilized to gain insight into molecular processes at lipid bilayer interfaces.<sup>33,34,49,50</sup> The template effect of vesicles for different polymerization reactions has been used to a great extent.<sup>51</sup> DOPC-templated tuning of product distribution in a dynamic combinatorial library has been reported in a few instances.<sup>41,42</sup> However, comparison of vesicles with different lipids in governing product distribution from a similar dimension of combinatorial library has not been explored. In this manuscript, as shown in Fig. 1 (bottom panel), we used nine different orders of the combinatorial matrix of aldehydes and hydrazines. We monitored the formation of hydrazones in

buffer, OA and DOPC vesicle environments using high-performance liquid chromatography (HPLC). In each case, we used 30  $\mu\text{M}$  of each reactant (both aldehyde and hydrazine); therefore, the maximum amount of product will also be 30  $\mu\text{M}$ , and we determined the amount of each individual product. Please see the SI for details of the experimental protocol.

We chose the aldehydes and hydrazine based on their partition coefficients and polar functionalities. We used the MarvinSketch tool from ChemAxon (MarvinJS), an online cheminformatics platform that allows for the prediction of  $\log D$  (distribution coefficient) in octanol–water systems at  $\text{pH} = 7$  (our experimental condition).<sup>41</sup> Each reactant (aldehydes and phenylhydrazine) was first drawn using the MarvinJS molecular editor, and then the  $\log D$  value was calculated, which provides an estimate of each molecule's relative hydrophobicity or hydrophilicity.  $\log D$  values of A1, A2 and A3 are 1.55, 1.63 and 1.98. We also checked benzaldehyde, which has a  $\log D$  value of 1.7, close to that of A2. Therefore, for our combinatorial library, we did not select benzaldehyde. Use of A2 instead of A4 will be beneficial, as our chosen A1 is 5-nitrosaldehyde. Addition of a nitro group will impart additional polarity to the molecule, and a comparison between A1 and A2 will generate information on an additional polar –OH group in the combinatorial hydrazone formation event. Additionally, the  $\log D$  values of H1, H2 and H3 were 1.2, 1.4 and 1.8, respectively. Therefore, for both aldehydes and hydrazines we chose the molecules with a wide-ranging (low to high) hydrophobicity and functional groups with varying degree of polarity for both aldehydes and hydrazines.

We also checked the encapsulation efficiency of the reagents in DOPC and OA vesicles by using an ultracentrifugation filtration experiment using a 3 kDa cut-off PES membrane (Fig. S1, SI).<sup>52,53</sup> With this method, we were able to find out the encapsulation of both aldehydes and hydrazines in DOPC vesicles, but unfortunately, not in OA vesicles. It is already mentioned in the literature that the native OA vesicle is unstable even in the presence of low-level salt ( $\sim 1 \text{ mM}$ ), and additionally, its membrane is highly permeable.<sup>30,48,54</sup> Therefore, during the ultracentrifugation experiment, there is a higher probability of leakage of the encapsulated molecules from their hydrophobic leaflet. Indeed, we find all the molecules in the filtrate, as nothing gets encapsulated in the OA vesicle. In case of the DOPC vesicle, almost 20–25  $\mu\text{M}$  reagents out of a maximum of 30  $\mu\text{M}$  (irrespective of aldehydes and hydrazines) were encapsulated inside the membrane after 1 h. We found encapsulated A1, A2 and A3 were  $20.5 \pm 0.8$ ,  $21 \pm 0.6$  and  $25 \pm 1.1 \mu\text{M}$ , respectively, and that is in order with the  $\log D$  value as discussed in the previous paragraph. Notably, A4 has a comparable encapsulation value ( $22 \pm 2 \mu\text{M}$ ) with 4-nitrobenzaldehyde (A2). In a similar way, we also checked the encapsulation of H1, H2 and H3 in the DOPC vesicle through the ultracentrifugation filtration experiment. The found values for H1, H2 and H3 were  $23 \pm 3$ ,  $23.5 \pm 1.4$  and  $25 \pm 2.4 \mu\text{M}$ , respectively. We want to note here that, irrespective of the low  $\log D$  value (1.2 and 1.4), we found reasonable uptake of H1 and H2 inside DOPC (more than A1 and A2, having higher  $\log D$  values) in our experimental protocol. Overall, all these experiments suggest more than 66–



90% encapsulation ability of our selected aldehydes and hydrazines in the DOPC vesicle membrane after 1 h.

Next, we used both transmission electron microscopy (TEM) and dynamic light scattering (DLS) to check the diameter of both types of vesicles, which were around 200–400 nm (Fig. S27–S29, SI). Furthermore, we used the general polarization value of Laurdan dye and found that OA vesicles are more gel-like (less hydrated) and DOPC vesicles are more liquid crystalline (more hydrated) in nature (Fig. S30, SI).<sup>55–57</sup>

At first, for analytical reference in HPLC, we synthesized all 9 hydrazones (Fig. S2–S26, SI). Next, we checked the hydrazone formation ability by simply using 1 aldehyde and 1 hydrazine, which can have a total of 9 individual combinations for a total of 3 aldehydes and 3 hydrazines. We checked the product formation after 1 h in each case. Here, we found that products containing polar moieties, A1H1 or A1H3, reached more than 70% yield in buffer or DOPC vesicles, considering their encapsulation efficiency. Interestingly, for A1H1, we found a  $24 \pm 1.3$   $\mu\text{M}$  product in the DOPC vesicle, which is slightly higher than its capacity formation considering individual encapsulation efficiency. This may be due to the release of the hydrazone product (A1H1) with polar groups ( $-\text{NO}_2$  and  $-\text{OH}$ ) to the aqueous environment from the hydrophobic leaflet of DOPC vesicles, allowing internalization of additional reactants. Interestingly, we found higher hydrophobic product (A3H3) formation ( $14 \pm 1.2$   $\mu\text{M}$ , close to 50% of maximum) in the OA vesicle in 1 h. Also, the amount of A2H3 (adduct with slightly polar aldehyde and non-polar hydrazine) was  $13 \pm 1.4$  and  $15 \pm 1.5$   $\mu\text{M}$ , respectively, in OA and DOPC vesicles after 1 h. Please check Table S1 in the SI for all the values in the  $1 \times 1$  case. As we found a very high amount of product formation in a few cases (and in one case complete conversion) after 1 h, we decided to fix this time point for our subsequent study with different higher-order matrices. Therefore, this time point can be treated as a reference to understand the amount of product formation in case of other higher-order matrices.

As per our planning, we followed to check all possible matrix order in a hierarchical way in terms of (number of aldehydes  $\times$  number of hydrazines) as following – ( $1 \times 2$ ), ( $1 \times 3$ ), ( $2 \times 1$ ), ( $2 \times 2$ ), ( $2 \times 3$ ), ( $3 \times 1$ ), ( $3 \times 2$ ) and ( $3 \times 3$ ). These experiments with hierarchical order, with all possible combinations, have been performed for the following reasons. Firstly, in this way all possible competition among substrates for product formation in aqueous buffer or gel-like OA and liquid-like DOPC-membranes can be generated when substrates (aldehydes or hydrazines) may or may not contain polar ( $-\text{NO}_2$  and  $-\text{OH}$ ), mildly polar ( $-\text{OCH}_3$ ) or non-polar groups ( $-\text{CH}_3$  or no specific functional group in the aromatic moiety). In this way, the environment (vesicle)-specific preference for polar or non-polar moieties with respect to buffer could be assessed, while simultaneously analyzing the influence of the reagent input ratios (matrix order) in directing the outcome of product distribution patterns. Secondly, as our aim was to find out the information entropy, the larger the data set, the more accurate the representation of the overall uncertainty and randomness within the data becomes. As mentioned in the preceding

paragraph, we checked the product distribution after 1 h in each case.

The amount of each product formed in all possible matrix orders has been tabulated in Fig. 2 and Tables S1–S17 (SI). It is to be noted that each product will appear 2, 2, 2, 2 and 4-times for matrix orders (no. of aldehydes  $\times$  no. of hydrazines) ( $1 \times 2$ ), ( $2 \times 1$ ), ( $3 \times 2$ ), ( $2 \times 3$ ) and ( $2 \times 2$ ) cases, respectively (see Tables S3–S19 in the SI for details). In Fig. 2, the average values of either 2 or 4 outcomes of each product have been given. For the remaining 4 matrix orders [( $1 \times 1$ ), ( $3 \times 1$ ), ( $1 \times 3$ ) and ( $3 \times 3$ )], each product will appear only once. The product formation data is tabulated in a colour-coded format. In the case of aqueous buffer, almost negligible amount of product formation was observed with A3 containing hydrazones and A2H2, whereas A1H1 and A1H3 formed in moderate to high amounts. This suggests that the presence of polar groups ( $-\text{NO}_2$  and  $-\text{OH}$ ) favours a higher amount of hydrazone formation in the buffer in competition. In the presence of OA vesicles, lower product formation occurred only in the case of H2-associated hydrazones, whereas in other cases, formation of hydrazones was noticeable. Strikingly, the formation of A3H3 in the OA vesicle was noticeably higher compared to the buffer and DOPC vesicle. For the OA vesicle, the amount of polar group containing hydrazone (with A1) is lower than in buffer, whereas weakly polar or non-polar group containing hydrazones (containing A2 and A3) were noticeably higher than in aqueous buffered media. In general, almost all the products (either lower or higher with respect to buffer) have been formed in the OA vesicle environment. Interestingly, in the DOPC vesicle, both polar and weakly polar groups containing hydrazone formation occurred. The highest amount of A1 containing hydrazones was found in the DOPC vesicular environment compared to both the buffer and OA vesicle. Additionally, A2 (with only a  $-\text{NO}_2$  group) containing hydrazone was formed in a higher amount in the DOPC vesicle (more than buffer but less than OA vesicles). A detailed quantitative assessment with fold enhancement in hydrazone formation for each matrix order in OA and DOPC vesicles compared to aqueous buffer has been provided in the SI (Fig. S33–S46, Tables S1–S17, SI).

Notably, from our GP value data, we found that the OA vesicle is gel-like, whereas the DOPC vesicle is liquid crystalline in nature. It is evident that non-polar molecules react efficiently in the OA vesicle, which is less-hydrated with only one C18 chain having carboxylic acid as the headgroup. In contrast, both polar and weakly polar hydrazones formed in DOPC vesicles inside their liquid-crystalline, more hydrated membrane. From our encapsulation study, we found that each aldehyde and hydrazine has almost equal encapsulation ability in the DOPC vesicle, at least when tested individually. However, the product formation is higher for polar or weakly polar hydrazones, indicating that higher membrane hydration and liquid crystalline nature may play a role in that. The permeability is higher for the OA vesicle (possibly due to the presence of only one hydrophobic tail, in contrast to two hydrophobic chains of the DOPC vesicle), which allowed both entrapping of reagents and release of products from their hydrophobic zone in a relatively





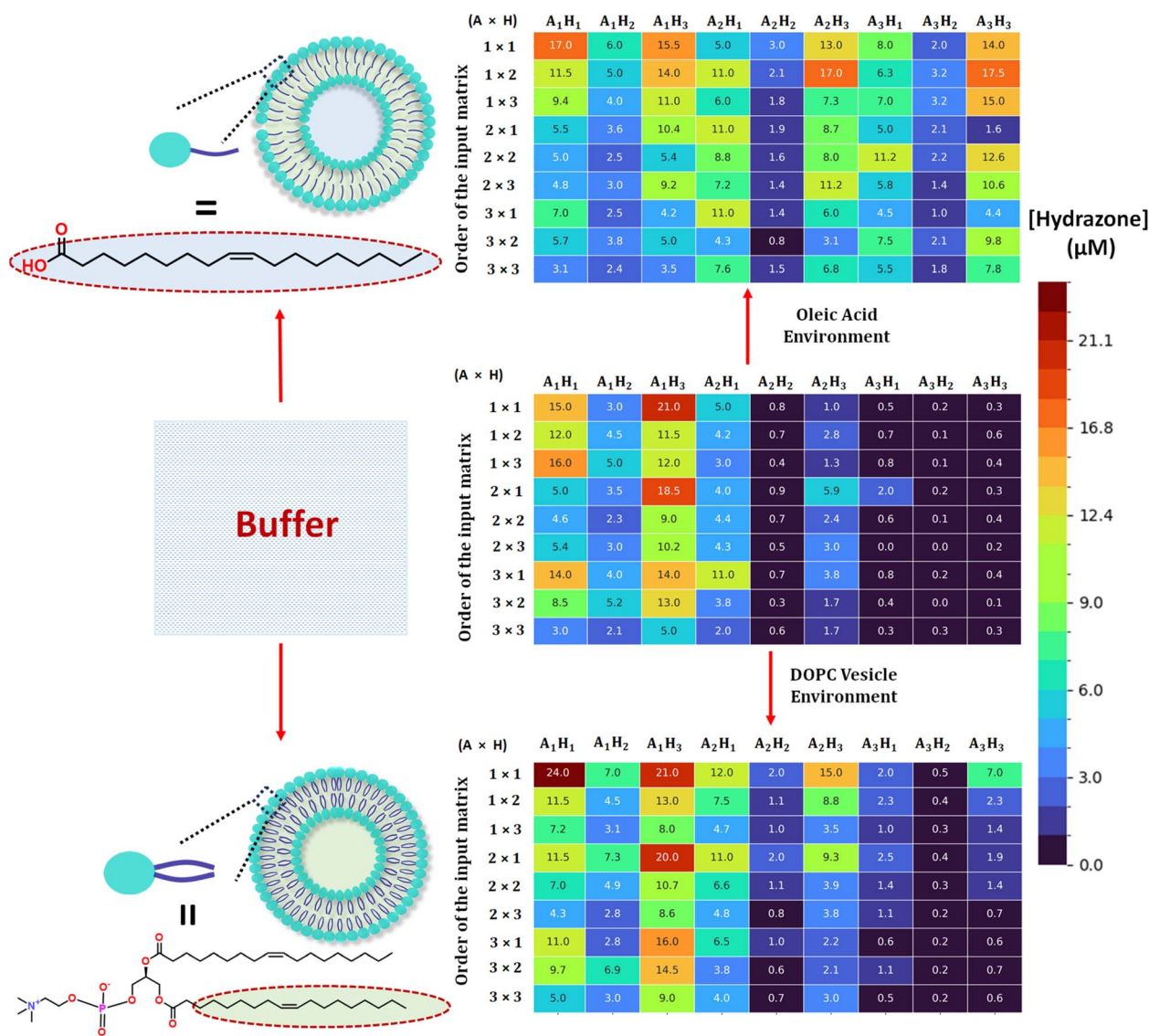


Fig. 2 Amount of product formation (concentration is depicted in the color code at the right) at different orders of the input matrix in aqueous buffer, DOPC and OA vesicles. Notably, DOPC contains two chains of the OA-based fatty acid part, designated by the red-dotted circle. Please see Tables S3–S19 and Fig. S33–S46 in the SI for detailed data and individual product concentration for certain input ratios ((1 × 2), (2 × 1), (3 × 2), (2 × 3) and (2 × 2) cases). Experimental conditions: [aldehyde] = 30 μM, [hydrazine] = 30 μM,  $T = 25^\circ\text{C}$ . All the experiments were carried out twice.

unselective manner.<sup>30,48,54</sup> This also led to a comparatively higher broad-spectrum hydrazone formation in the OA vesicle.

The main aim of this work is to quantify the information entropy (Shannon entropy) as well as other statistical parameters, such as the Gini coefficient. In this case, Shannon entropy and the Gini coefficient were used to analyze the nature of the environment and ratios. In other words, we sought to understand what sort of products are likely to form under different conditions of environment and initial reactant ratios. Each data point is grouped by ratio and environment, and the mean concentration of each group is taken as the data point for this analysis. In particular, Shannon entropy measures diversity or uncertainty. In principle, in our case, this can be used for quantifying how many products are formed and how evenly they

are distributed. High Shannon entropy would mean there is higher uncertainty about which product will be found/formed and low entropy would mean only a few products dominate, and therefore, there is lower uncertainty.

Shannon entropy is given by the formula:

$$H = -\sum_i p_i \log_2 p_i$$

where  $p_i$  is the normalized concentration of product  $i$ .

The Gini coefficient measures inequality in a distribution. Originally used in economics to measure income inequality, it is also used in chemistry to quantify selectivity – in this case, how much a certain environment favors one product over others.<sup>58</sup> The resulting value ranges from 0 to 1. A Gini coefficient value of



0 indicates that all products are present in exactly equal amounts and that there is no selectivity (perfect equality), while a Gini coefficient of 1 indicates that only one product is present and all others are zero, showing perfect selectivity (maximum inequality).

The Gini coefficient is calculated using:

$$G = \frac{\sum_{i=1}^n \sum_{j=1}^i |x_i - x_j|}{2n \sum_{i=1}^n x_i}$$

where  $x_i$  is the concentration of product  $i$  and  $n$  is the number of products.

At first, we plotted Shannon entropy (SE) (on the X-axis) and Gini coefficient (GC) values (on the Y-axis) for each hydrazone product in all three different environments (aqueous buffer, OA and DOPC vesicles) when introduced at 9 different matrix orders (Fig. 3a). In the case of aqueous buffer, the highest GC (around 0.3) was observed for A1H1, A2H3 and A3H1, showing their formation is selective to the input matrix order. In buffer, the lowest SE (signifying least diversity with respect to input ratios) was observed for A3H2, as the formation of this product was almost negligible. A high SE value was observed for A1H2, A1H3 and A2H2, showing that the formation probability of these products is more diverse; in other words, predicting these three-product formation with respect to input ratio was more uncertain. Interestingly, the highest GC value of 0.46 and the lowest SE value of 2.6 were observed for A3H3 in the DOPC vesicle. It is due

to the selective formation of A3H3 at the  $(1 \times 1)$ -matrix case. Additionally, a considerably higher GC value was observed for A2H3, showing its selective formation in certain input ratios with lower competition. Interestingly, the lowest GC value and highest SE value were obtained for A1H3 in the DOPC vesicle. In contrast to the DOPC vesicle, for the OA vesicle, the highest GC value and moderate SE were observed for A1H3, indicating that the probability of this product formation is restricted to a few low-order matrices (specifically,  $1 \times 1$ ). After general analysis, it can be predicted that the SE value of higher than 3.0 was obtained for 4, 5 and 6 hydrazones in aqueous buffer, DOPC and OA vesicles, respectively, while varying the order of the input matrix. This suggests that the order of uncertainty or SE followed the order: OA vesicle > DOPC vesicle > aqueous buffer.

Next, we plotted the GC vs. SE values in terms of nine hydrazone products formation diversity at each order of the input matrix (Fig. 3b). This will indicate the formation of selective one or a few products or more products in an unselective manner in different environments. The data clearly suggest a low GC (0.3–0.36) and high SE (2.85–3.0) in the case of the OA vesicle. A moderate GC (0.5, 0.56 and 0.59 for each of  $(3 \times 3)$ ,  $(1 \times 3)$  and  $(3 \times 1)$  cases and 0.43–0.46 for the remaining 6 cases) and a moderate SE value (2.3–2.7) were observed in the DOPC vesicle. Again, in aqueous buffer, a comparatively lower SE and higher GC value were found. These data again clearly indicate that the uncertainty or probability of diversified product formation is highest for the OA vesicle, followed by the DOPC vesicle and aqueous buffer.

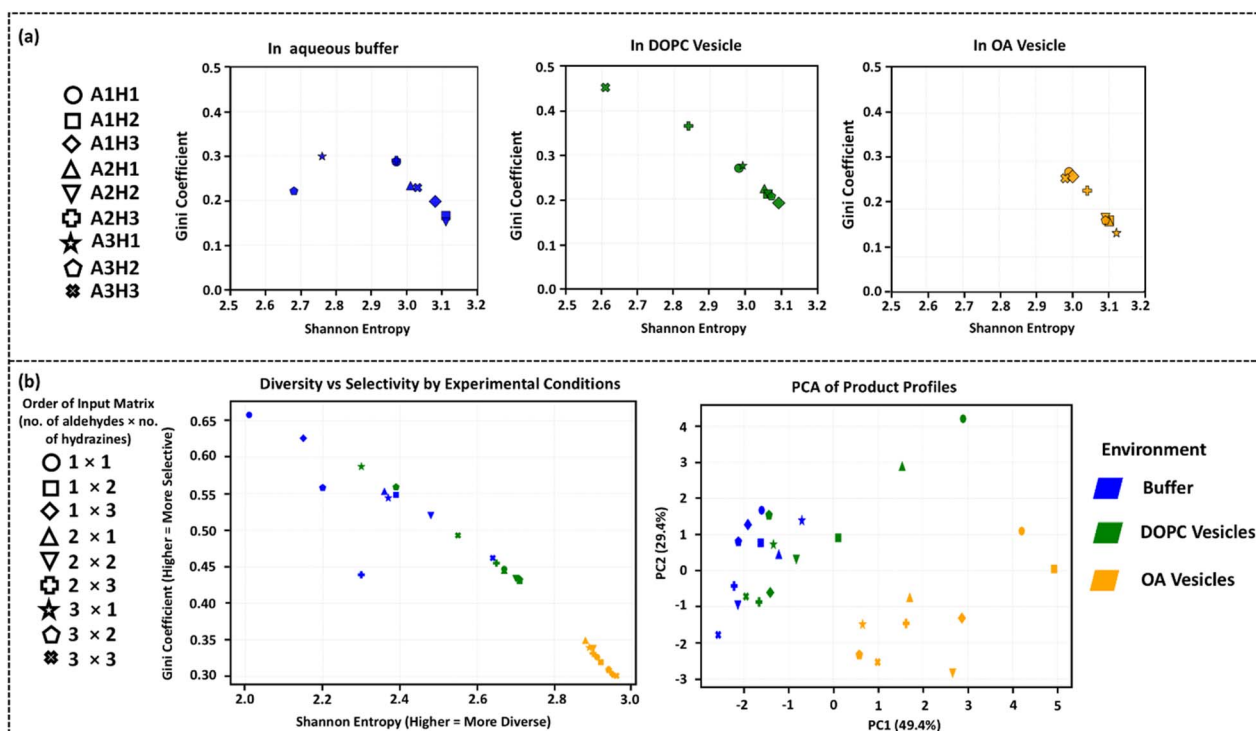


Fig. 3 (a) Plot of GC vs. SE depicting each hydrazone (represented by different symbols) in aqueous buffer, DOPC and OA vesicles for varying input ratios or orders of the matrix. (b) Left panel depicts GC vs. SE plot of each order of the input matrix when 9-hydrazones products are variable. Right panel depicts the principal component analysis plot to show the variance in the dataset and clustering of OA-vesicle specific datapoints at a distal zone from the buffer and DOPC vesicle.



Table 1 Interaction between the input ratio and environment: two-way ANOVA

Product	Is ratio significant?	Is environment significant?	Is their interaction significant?	Interpretation
A1H1	Yes ( $p = 5.8 \times 10^{-9}$ )	Yes ( $p = 0.0058$ )	Yes ( $p = 0.0136$ )	Both ratio and environment matter, and their interaction affects yields
A1H2	Yes ( $p = 0.0014$ )	Yes ( $p = 0.0009$ )	No ( $p = 0.0627$ )	Ratio and environment act independently; OA 2 : 1 is lower than both buffer and DOPC
A1H3	Yes ( $p = 0.00085$ )	Yes ( $p = 0.0029$ )	No ( $p = 0.535$ )	Ratio dominates; environment has a smaller effect
A2H1	No ( $p = 0.064$ )	Yes ( $p = 0.0028$ )	No ( $p = 0.544$ )	Environment alone drives concentration
A2H2	Yes ( $p = 0.0035$ )	Yes ( $p = 3.1 \times 10^{-6}$ )	No ( $p = 0.632$ )	Strong environmental effect, moderate ratio effect
A2H3	Yes ( $p = 0.0002$ )	Yes ( $p = 5.0 \times 10^{-7}$ )	Yes ( $p = 0.027$ )	Both factors and interaction matter; buffer 2 : 3 has very low concentration
A3H1	No ( $p = 0.380$ )	Yes ( $p = 1.5 \times 10^{-9}$ )	No ( $p = 0.338$ )	Environment is sole driver; OA 2 : 1 shows notable amplification
A3H2	No ( $p = 0.336$ )	Yes ( $p = 1.3\text{e-}10$ )	No ( $p = 0.559$ )	Extremely strong environmental effect
A3H3	No ( $p = 0.088$ )	Yes ( $p = 6.0 \times 10^{-9}$ )	No ( $p = 0.172$ )	Environment dominates; OA 2 : 1 is much lower compared to other ratios in the same environment

Furthermore, we used Principal Component Analysis (PCA), which is a statistical technique used to reduce the dimensionality of datasets while retaining the most important components that explain the variability in the data. For our analysis, we measure the mean concentration of each product under each reactant ratio and environment pair. PC1 and PC2 are the two principal axes, which represent the linear combinations of product concentrations that explain the most variance. The PCA function from the Python library 'sklearn' was used to plot, and the results are displayed in Fig. 3c. PC1 and PC2 together can explain 78% of the total variance in the dataset. The clustering of data indicates that the reaction environment has a strong effect on product distribution. It also clearly suggests that the environment in the OA vesicle is significantly different from that in the DOPC vesicle and buffer. In fact, the environment for the DOPC vesicle and buffer can also be differentiated in this analysis.

Finally, we used two-way ANOVA, which is a statistical method to determine how two independent categorical variables (in this case, environment and input ratio or order of the matrix) and their interactions affect a dependent variable (product concentration) (Tables 1 and S18, SI). It tests main effects (the effect of each variable separately) and mixed effects (whether the effect of one factor depends on the level of the other).

The formula used is:

$$\text{Concentration} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon$$

where  $\mu$  is the overall mean concentration,  $\alpha_i$  is the effect of the  $i^{\text{th}}$  ratio,  $\beta_j$  is the effect of the  $j^{\text{th}}$  environment and  $\varepsilon$  is the random error.

Reviewing the product distributions across environments and ratios points to some noteworthy outliers. The ratio effect dominates the environment in the case of A1H3; however, in the OA environment at a 2 : 1 reactant combination, the product yield drops to roughly half its value compared to other environments at the same ratio. In the case of A3H3, where the systems-level analysis concludes that the environmental effect dominates, a surprisingly low concentration is observed in the case of the OA environment with a 2 : 1 combination ratio. Indeed, we found lower formation of the A3-product in the OA vesicle for both 2 : 1 and 3 : 1 combinations, in comparison to 1 : 2 or 1 : 3 combinations. Specifically, formation of A3H3 is strikingly low (only 0.2  $\mu\text{M}$ ) for the 2 : 1 case when A2 and A3 were added with H3 (Table S4, SI). In this case, the amount of A2H3 was much higher ( $6.4 \pm 0.7 \mu\text{M}$ ). This clearly suggests that the preference of hydrazine in the presence of multiple aldehydes is very different from its other counterpart (hydrazone formation propensity of one aldehyde in the presence of multiple hydrazines) in the OA vesicle. Overall, all these data clearly indicate the importance of the ratio and combination of inputs towards the outcome of products in different environments.

## Conclusions

At first, we separately showed the template effect of a single chain fatty acid (OA) vesicle and a double chain (comprising the same OA unit) phospholipid (DOPC) vesicle towards hydrazone product distribution through hierarchical input matrix ordering of aldehyde and hydrazine. Through detailed analysis of all possible types of competing environments between polar and non-polar



groups containing aldehydes and hydrazines, we extracted important statistical parameters (SE and GC) related to information theory. Additionally, *via* the two-way Anova method, the importance of input ratio and vesicle-specific environment for the emergence of each hydrazone product (with polar, mildly polar and non-polar moieties) has been interpreted. Indeed, we found the highest information (Shannon) entropy for the OA vesicle, signifying high diversity and low specificity; whereas the non-templated aqueous buffered system showed low Shannon entropy and a high Gini coefficient, signifying limited variation in distribution. Interestingly, the information storage capacity of the DOPC vesicle lies between that of the OA vesicle and aqueous buffer. Nevertheless, considering OA (fatty acid) vesicles appeared prior to DOPC vesicles on early Earth, the higher information entropy in OA vesicles compared to DOPC indicates the higher degree of diversity may be persisted in the prebiotic world. Simultaneously, a higher selectivity (GC value) in DOPC liposomes compared to OA vesicles delineates that complexification of lipids eventually helped in imposing order in terms of molecular uptake and reactivity in an organized compartmentalized system.

In this work, we carried out all the experiments at a single time point, as our main goal was to realize the information entropy *via* calculating SE and GC values. However, it is well-known that hydrazones can dissociate (hydrolyze back to aldehyde and hydrazine) in aqueous media.<sup>59,60</sup> Considering the higher permeability of the OA vesicle compared to the DOPC vesicle, it will be interesting to check the time evolution of the dynamic library for a longer time period. Indeed, the time-evolution study will give us the kinetic and thermodynamic scenarios of both the vesicle, which is the long-term goal. At this point, we are exploring this temporal aspect with different fatty acids (having varying no. of double bonds from 0–3), phospholipids, and their mixtures.

Additionally, this mode of quantification can indeed be utilized for different liposomes containing multiple lipids as well as other ions, amino acids or nucleotides, including compositions, to understand the progress of information in each system, which is relevant for the lipid world hypothesis.<sup>61–63</sup> We believe this work opens up the possibility of using self-assembled systems – such as liposomes, micelles – as information storage devices similar to nucleic acids and small organic molecules.<sup>64–66</sup> Indeed, apart from the vesicle-templated system, future work will aim to delineate information entropy for nanoparticle- or different self-assembled monolayer-based systems using both dynamic and dissipative-dynamic combinatorial chemistry approaches as a function of time.<sup>67–73</sup> This might help to identify information memory or how a system quantitatively learn its previous information footprints.<sup>74</sup> We believe that the potential of this research will indeed be far-fetched – for example, in understanding (and quantifying) the information dynamicity of different chemically fuelled or self-propagating synthetic and natural assembly processes over time.

## Author contributions

RY performed most of the experiments and data acquisition. NA did all the statistical analysis. RRM performed part of the

experiments. SM designed the experiment, supervised the work, and wrote the manuscript. All authors commented on the manuscript.

## Conflicts of interest

There are no conflicts to declare.

## Data availability

The data supporting this article have been included as part of the SI. Complete synthetic procedures, additional NMR, HPLC, TEM, and fluorescence spectroscopy data are provided in the SI. See DOI: <https://doi.org/10.1039/d5sc04365d>.

## Acknowledgements

S. M. acknowledges financial support of ANRF (File No. CRG/2022/002345). RY and RRM acknowledge IISER Mohali for the doctoral research grant. We also acknowledge DST-FIST for the 400 MHz NMR-facility in the Department of Chemical Sciences of IISER Mohali.

## Notes and references

- 1 J. Machta, *Am. J. Phys.*, 1999, **67**, 1074–1077.
- 2 I. Ben-Gal and E. Kagan, *Entropy*, 2021, **23**, 232.
- 3 C. Adami, *Phys. Life Rev.*, 2004, **1**, 3–22.
- 4 A. Golan and J. Harte, *Proc. Natl. Acad. Sci. U. S. A.*, 2022, **119**, e2119089119.
- 5 C. Shannon, in *Ideas That Created the Future*, The MIT Press, 2021, pp. 121–134.
- 6 L. Szilard, *Syst. Res.*, 1964, **9**, 301–310.
- 7 H. P. Yockey, *Inf. Sci.*, 2002, **141**, 219–225.
- 8 B.-O. Küppers, *Information and the origin of life*, 1990, <https://mitpress.mit.edu/9780262111423/information-and-the-origin-of-life>.
- 9 S. Chirumbolo and A. Vella, *Molecules*, 2021, **26**, 1003.
- 10 B. J. Cafferty, A. S. Ten, M. J. Fink, S. Morey, D. J. Preston, M. Mrksich and G. M. Whitesides, *ACS Cent. Sci.*, 2019, **5**, 911–916.
- 11 N. Mac Fhionnlaoich and S. Guldin, *Chem. Mater.*, 2020, **32**, 3701–3706.
- 12 D. S. Sabirov and I. S. Shepelevich, *Entropy*, 2021, **23**, 1240.
- 13 G. Karreman, *Bull. Math. Biophys.*, 1955, **17**, 279–285.
- 14 D. S. Sabirov, *Comput. Theor. Chem.*, 2018, **1123**, 169–179.
- 15 D. Sh. Sabirov, *Comput. Theor. Chem.*, 2020, **1187**, 112933.
- 16 M. Soete, C. Mertens, N. Badi and F. E. Du Prez, *J. Am. Chem. Soc.*, 2022, **144**, 22378–22390.
- 17 I. Paul, I. Valiyev and M. Schmittel, *J. Am. Chem. Soc.*, 2024, **146**, 2435–2444.
- 18 S. L. Rössler, N. M. Grob, S. L. Buchwald and B. L. Pentelute, *Science*, 2023, **379**, 939–945.
- 19 N. F. König, A. Al Ouahabi, L. Oswald, R. Szweda, L. Charles and J.-F. Lutz, *Nat. Commun.*, 2019, **10**, 3774.
- 20 J.-M. Lehn, *Angew. Chem., Int. Ed.*, 2013, **52**, 2836–2850.
- 21 J.-M. Lehn, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 4763–4768.





- 22 P. L. Gentili and P. Stano, *ChemSystemsChem*, 2024, **6**, e202400054.
- 23 P. L. Gentili, *Molecules*, 2020, **25**, 3634.
- 24 R. Landauer, *Phys. Today*, 1991, **44**, 23–29.
- 25 T. Rieu, A. Osypenko and J.-M. Lehn, *J. Am. Chem. Soc.*, 2024, **146**, 9096–9111.
- 26 C. Gini, *Variabilità e mutabilità, In Studio delle Distribuzioni e delle Relazioni Statistiche*, ed. C. Cuppini, Bologna, 1912.
- 27 A. Ursu, J. L. Childs-Disney, A. J. Angelbello, M. G. Costales, S. M. Meyer and M. D. Disney, *ACS Chem. Biol.*, 2020, **15**, 2031–2040.
- 28 I. E. Weidlich and I. V. Filippov, *J. Comput. Chem.*, 2016, **37**, 2091–2097.
- 29 S. O'Hagan, M. Wright Muelas, P. J. Day, E. Lundberg and D. B. Kell, *Cell Syst.*, 2018, **6**, 230–244.
- 30 L. Jin, N. P. Kamat, S. Jena and J. W. Szostak, *Small*, 2018, **14**, e1704077.
- 31 K. Adamala and J. W. Szostak, *Science*, 2013, **342**, 1098–1100.
- 32 P. Walde, *Bioessays*, 2010, **32**, 296–303.
- 33 J. W. Hindley, Y. Elani, C. M. McGilvery, S. Ali, C. L. Bevan, R. V. Law and O. Ces, *Nat. Commun.*, 2018, **9**, 1093.
- 34 S. Patra, S. Dhiman and S. J. George, *Angew. Chem., Int. Ed.*, 2025, **64**, e202500456.
- 35 D. Segrè, D. Ben-Eli, D. W. Deamer and D. Lancet, *Origins Life Evol. Biospheres*, 2001, **31**, 119–145.
- 36 D. Lancet, D. Segrè and A. Kahana, *Life*, 2019, **9**, 77.
- 37 V. Subbotin and G. Fiksel, *Astrobiology*, 2023, **23**, 344–357.
- 38 S. Pulletikurti, K. S. Veena, M. Yadav, A. A. Deniz and R. Krishnamurthy, *Chem*, 2024, **10**, 1839–1867.
- 39 T. C. B. Santos and A. H. Futerman, *Prog. Lipid Res.*, 2023, **92**, 101253.
- 40 K. A. Podolsky and N. K. Devaraj, *Nat. Rev. Chem.*, 2021, **5**, 676–694.
- 41 C. Bravin and C. A. Hunter, *Chem. Sci.*, 2020, **11**, 9122–9125.
- 42 C. Bravin, N. Duindam and C. A. Hunter, *Chem. Sci.*, 2021, **12**, 14059–14064.
- 43 S. Maiti, I. Fortunati, C. Ferrante, P. Scrimin and L. J. Prins, *Nat. Chem.*, 2016, **8**, 725–731.
- 44 Priyanka, S. Kaur Brar and S. Maiti, *ChemNanoMat*, 2022, **8**, e202100498.
- 45 M. A. Cardona and L. J. Prins, *Chem. Sci.*, 2019, **11**, 1518–1522.
- 46 S. Chandrabhas, S. Maiti, I. Fortunati, C. Ferrante, L. Gabrielli and L. J. Prins, *Angew. Chem., Int. Ed.*, 2020, **59**, 22223–22229.
- 47 A. Kamra, S. Das, P. Bhatt, M. Solra, T. Maity and S. Rana, *Chem. Sci.*, 2023, **14**, 9267–9282.
- 48 A. Rendón, D. G. Carton, J. Sot, M. García-Pacios, L.-R. Montes, M. Valle, J.-L. R. Arrondo, F. M. Goñi and K. Ruiz-Mirazo, *Biophys. J.*, 2012, **102**, 278–286.
- 49 M. De Poli, W. Zawodny, O. Quinonero, M. Lorch, S. J. Webb and J. Clayden, *Science*, 2016, **352**, 575–580.
- 50 M. A. Watson and S. L. Cockroft, *Chem. Soc. Rev.*, 2016, **45**, 6118–6129.
- 51 A. M. van Herk, *Biomacromolecules*, 2020, **21**, 4379–4387.
- 52 Priyanka and S. Maiti, *Langmuir*, 2024, **40**, 18906–18916.
- 53 N. K. Ettikkan, P. Priyanka, R. R. Mahato and S. Maiti, *Commun. Chem.*, 2024, **7**, 242.
- 54 C. L. Apel, D. W. Deamer and M. N. Mautner, *Biochim. Biophys. Acta*, 2002, **1559**, 1–9.
- 55 R. Yadav, N. Sivoria and S. Maiti, *J. Phys. Chem. B*, 2024, **128**, 9573–9585.
- 56 Priyanka and S. Maiti, *J. Mater. Chem. B*, 2023, **11**, 10383–10394.
- 57 A. Deshwal and S. Maiti, *Langmuir*, 2021, **37**, 7273–7284.
- 58 A. Nishi, H. Shirado, D. G. Rand and N. A. Christakis, *Nature*, 2015, **526**, 426–429.
- 59 E. Potůčková, K. Hrušková, J. Bureš, P. Kovaříková, I. A. Špírková, K. Pravidíková, L. Kolbabová, T. Hergeselová, P. Hašková, H. Jansová, M. Macháček, A. Jirkovská, V. Richardson, D. J. R. Lane, D. S. Kalinowski, D. R. Richardson, K. Vávrová and T. Šimůnek, *PLoS One*, 2014, **9**, e112059.
- 60 J. Kalia and R. T. Raines, *Angew. Chem., Int. Ed.*, 2008, **47**, 7523–7526.
- 61 D. Lancet, R. Zidovetzki and O. Markovitch, *J. R. Soc. Interface*, 2018, **15**, 20180159.
- 62 J. Pereto, *Chem. Soc. Rev.*, 2012, **41**, 5394–5403.
- 63 M. Fiore, C. Chieffo, A. Lopez, D. Fayolle, J. Ruiz, L. Soulère, P. Oger, E. Altamura, F. Popowycz and R. Buchet, *Astrobiology*, 2022, **22**, 598–627.
- 64 L. C. Meiser, B. H. Nguyen, Y.-J. Chen, J. Nivala, K. Strauss, L. Ceze and R. N. Grass, *Nat. Commun.*, 2022, **13**, 352.
- 65 B. Liu, F. Wang, C. Fan and Q. Li, *Adv. Mater.*, 2025, **37**, 2314358.
- 66 A. A. Nagarkar, S. E. Root, M. J. Fink, A. S. Ten, B. J. Cafferty, D. S. Richardson, M. Mrksich and G. M. Whitesides, *ACS Cent. Sci.*, 2021, **7**, 1728–1735.
- 67 P. T. Corbett, J. Leclaire, L. Vial, K. R. West, J.-L. Wietor, J. K. M. Sanders and S. Otto, *Chem. Rev.*, 2006, **106**, 3652–3711.
- 68 R. A. R. Hunt and S. Otto, *Chem. Commun.*, 2011, **47**, 847–858.
- 69 S. Maiti and L. J. Prins, *Chem. Commun.*, 2015, **51**, 5714–5716.
- 70 A. Osypenko, R. Cabot, J. J. Armao IV, P. Kovaříček, A. Santoro and J.-M. Lehn, *ChemistryEurope*, 2023, **1**, e202300017.
- 71 K. Das, L. Gabrielli and L. J. Prins, *Angew. Chem., Int. Ed.*, 2021, **60**, 20120–20143.
- 72 D. Del Giudice, M. Valentini, G. Melchiorre, E. Spatola and S. Di Stefano, *Chem.–Eur. J.*, 2022, **28**, e202200685.
- 73 C. M. E. Kriebisch, L. Burger, O. Zozulia, M. Stasi, A. Floroni, D. Braun, U. Gerland and J. Boekhoven, *Nat. Chem.*, 2024, **16**, 1240–1249.
- 74 K. Kaygisiz and R. V. Ulijn, *ChemSystemsChem*, 2025, **7**, e202400075.

