



Soft Matter

**In search of a novel chassis material for synthetic cells:
emergence of synthetic peptide compartment**

Journal:	<i>Soft Matter</i>
Manuscript ID	SM-REV-09-2020-001644.R1
Article Type:	Review Article
Date Submitted by the Author:	29-Oct-2020
Complete List of Authors:	Sharma, Bineet; University of Michigan, Mechanical Engineering Ma, Yutao; University of Chicago, Institute for Molecular Engineering Ferguson, Andrew; University of Chicago, Institute for Molecular Engineering Liu, Allen; University of Michigan, Mechanical Engineering

SCHOLARONE™
Manuscripts

In search of a novel chassis material for synthetic cells: emergence of synthetic peptide compartment

Bineet Sharma¹, Yutao Ma², Andrew L Ferguson², Allen P Liu^{1,3,4,5}

¹ Department of Mechanical Engineering, University of Michigan, Ann Arbor, Michigan 48109 USA

² Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 60637, USA

³ Department of Biomedical Engineering, University of Michigan, Ann Arbor, Michigan 48109, USA

⁴ Cellular and Molecular Biology Program, University of Michigan, Ann Arbor, Michigan 48109, USA

⁵ Department of Biophysics, University of Michigan, Ann Arbor, Michigan 48109, USA

Co-corresponding authors: ALF: andrewferguson@uchicago.edu (ORCID: 0000-0002-8829-9726), APL: allenliu@umich.edu (ORCID: 0000-0002-8829-9726)

Abstract

Giant lipid vesicles have been used extensively as a synthetic cell model to recapitulate various life-like processes, including *in vitro* protein synthesis, DNA replication, and cytoskeleton organization. Cell-sized lipid vesicles are mechanically fragile in nature and prone to rupture due to osmotic stress, which limits their usability. Recently, peptide vesicles have been introduced as a synthetic cell model that would potentially overcome the aforementioned limitations. Peptide vesicles are robust, reasonably more stable than lipid vesicles and can withstand harsh conditions including pH, thermal, and osmotic variations. This mini-review summarizes the current state-of-the-art in the design, engineering, and realization of peptide-based chassis materials, including both experimental and computational work. We present an outlook for simulation-aided and data-driven design and experimental realization of engineered and multifunctional synthetic cells.

1. Introduction

The cell, the "simplest" unit of life, is in fact a highly sophisticated system. It is a multi-compartment system that houses vital reactions that are coordinated spatiotemporally between different organelles. In the course of understanding life, it is essential to comprehend how a micron-sized cell is capable of organizing all the necessary components for its survival. A simple start towards answering this question would be to create the basic structure of the cell. This has guided the concept of synthetic cells in bottom-up synthetic biology whereby different processes including protein synthesis, cell division, and endocytosis-exocytosis are reconstituted in a cell-like system.¹⁻⁷

The first step in building a synthetic cell system is the creation of a compartment. Compartmentalization enables the encapsulation of DNAs, RNAs, and proteins providing a wealth of applications.⁸⁻¹⁰ The compartments can vary in sizes from few hundred nanometers to few hundred micrometers and can be prepared by numerous methods such as thin film hydration method, electroformation, tetrahydrofuran (THF) swelling, and extrusion. Compartments formed by these methods are termed as vesicles, often used as a synthetic cell-like model where the exchange of the materials from inner solution to outer solution and vice-versa is limited. Vesicles are composed of a bilayer which mimics the biological cell membrane. The bilayer membrane is generated through the self-assembly of lipid molecules containing a hydrophilic head and a hydrophobic tail. Lipid vesicles can be prepared by using different methods yielding sizes ranging from nanometers to tens of micrometers in diameter. These vesicles are extensively employed in different fields from synthetic biology to applied sciences like nanotechnology, drug delivery and diagnostics.^{11,12}

Aside from compartmentalizing biochemical reactions, the cell membrane itself mediates many vital reactions such as cell-cell communication, redox reactions at mitochondrial membrane, endocytosis/exocytosis, and signal transmission in a nerve cell. In the context of deciphering the roles of the membrane, it is necessary to consider both the biochemical and biophysical aspects of a bilayer membrane. Conventional lipid bilayer vesicles have high inherent deformability than natural cells as model membrane vesicles do not have a supporting cytoskeleton network. Hence, lipid vesicles are susceptible to mechanical stress and are also susceptible to harsh conditions such as high salt concentration. The fragile nature of cell-sized lipid vesicles has motivated the exploration of alternative chassis material for synthetic cells. Chemically synthesized amphiphilic polymers were introduced as polymersomes that have immense biomedical industrial applications¹³ and have been used in the fabrication of hybrid vesicles consisting of lipid and polymer to provide enhanced mechanical strength.¹⁴ These

hybrid vesicles, made of lipid-poly(butadiene)-b-poly(ethylene oxide), have been shown to provide a better folding environment for a mechanosensitive channel in the membrane.¹⁵ Nevertheless, polymersomes are not fully biocompatible and this has restricted their usability.

Recently, amphiphilic peptide vesicles based on elastin-like polypeptides (ELPs) were introduced by Vogele et al. as an artificial cell chassis material.¹⁶ Peptide vesicles are mechanically robust and can survive high thermal, osmotic and chemical assaults. Inspired by these recent advances and the potential for using amphiphilic peptides as a building block for synthetic cells, we will summarize recent literature on different types of peptide-based vesicles, highlighting the advantages and associated challenges. We will also review computational studies that aim to understand and engineer these polypeptides. Finally, we provide some outlook of using ELP vesicles for bottom-up synthetic biology studies.

2. Self-assembly of amphiphilic polypeptide prototypes: Distinct peptidic scaffolds

2.1 Natural amphiphilic polypeptides

Surface active proteins, or surfactant proteins, are typically amphiphilic polypeptides that lower the surface tension at the water-air (mostly) and water-oil interfaces. This effect is crucial for many life processes including sporulation in fungi and bacteria, formation of foam nests in frog reproduction, and cooling in horse's sweat and saliva (see **Table 1**). These naturally-occurring biological surfactants have been repurposed for a variety of applications from coating of nano-devices and medical implants to food industry as emulsifier and in personal care products.^{17,18} We briefly introduce a number of key natural amphiphilic polypeptides below and refer interested readers to a detailed review by Schor et al.¹⁹

Oil bodies present in plant seeds play multiple roles in plant physiology such as lipid metabolism and hormonal signaling.²⁰ They also serve as reservoirs of toxic fatty acids that are lethal to plant cells. These oil bodies are stabilized by surfactant proteins called oleosins.^{21,22} Oleosins are a class of plant proteins of size 15 – 26 kDa, with three different regions, N-terminal hydrophilic – middle hydrophobic – C-terminal hydrophilic regions.^{23,24} Oleosin has one of the longest natural hydrophobic domains.²⁵ Self-assembly of recombinant oleosins was reported by the Hammer group (see **Figure 1A**) where they investigated the phase behavior of different oleosins as a function of ionic strength and the hydrophilic length at N- and C- termini while keeping the middle hydrophobic region unchanged. The formation of giant protein bilayer vesicles from phase separation of double emulsions was demonstrated, highlighting the use of recombinant amphiphilic proteins to self-assemble into suprastructures.²⁶

Hydrophobins are a class of small globular surface-active proteins of ~75-150 amino acid residues. They are one of the most studied and applied surface-active proteins. Hydrophobins are produced by filamentous fungi *Schizophyllum commune* and play an important role in fungal growth and development.²⁷ The submerged part of this fungi secretes hydrophobins which self-assemble at air-water interface, lower the surface tension, and assist in the aerial growth of the fungi. Structurally, hydrophobins are amphiphilic in nature with a hydrophilic region and a “hydrophobic patch” comprised of aliphatic residues (see **Figure 1B**). They have eight highly conserved cysteine residues which form four intramolecular disulfide bonds.^{28,29} Hähl et al. reported a microfluidic approach for making free-standing protein bilayers from native hydrophobins. They showed the assembly of hydrophobins with either a hydrophobic core or a hydrophilic core as the bilayer as shown in **Figure 1B**, and demonstrated giant protein vesicle formation by using microfluidic jetting.³⁰

Ranaspumins comprise another class of important surface-active proteins. They are important stabilizing proteins in organisms where external fertilization occurs, mostly in fish and frogs. Before fertilization, a male frog generates a foam nest by vigorously peddling using a small amount of surface-active liquid released by a female frog. The air bubble nest generated at the air-water interface with ranaspumins and other stabilizing agents help the fertilized eggs develop into tadpoles. The foam nest does not coalesce due to the presence of ranaspumins at the interface and is stable for weeks. It also protects the fertilized eggs from dehydration and microbes.^{31,17} Ranaspumins are monomeric in nature with a size up to 25 kDa and their high-resolution x-ray structure and solution NMR structure have been determined.^{32,33} Interestingly, it is believed that they undergo conformational changes to reduce the interfacial tension along with other stabilizing agents.³² As an example, artificial photosynthetic fixation of CO₂ into sugar and light into ATP are demonstrated by Wendell et al. using bacteriorhodopsin/F₀F₁ ATPase lipid vesicles.³⁴ The lipid vesicles are trapped at the interface of the foam stabilized by Ranaspumin-2.

An animal-based surfactant protein called latherin is a non-glycosylated surface-active proteins present in sweat and saliva of horses and equids.^{35,36} Unlike humans, a horse's sweat has low salt and higher protein content. A major part of these proteins is composed of latherin, and like other surfactant proteins it also decreases the interfacial tension. It acts as a wetting agent and facilitates the evaporation from the surface of the horse's skin and regulates the body temperature. It also prevents deposition of microbial biofilm in the pelt.³⁷ Latherin is composed of a high fraction of hydrophobic amino acids and has exceptionally high leucine content.³⁵ Like ranaspumins, it is present in monomeric state and has no amphiphilic nature. In the absence of

any amphiphilic character, a mechanism of action in lowering surface tension has been proposed based on an interfacial conformational change.³⁸

2.2 Synthetic polypeptides

2.2.1 Chemically synthesized polypeptides

Block co-polypeptides are the crosslinked product of two or more polypeptides. These polypeptides are extensively utilized in the field of biomaterials.³⁹ As shown in **Figure 2A**, polypeptides are chemically synthesized by acid-N-carboxyanhydride monomers of amino acids and combined via various methods such as alkyne-azide conjugation or crosslinked by oxidation or by using a crosslinker.^{39,40} Chemically synthesized polypeptides are relatively easy to formulate di-, tri or multi-block co-polypeptides with other existing methods. The chemical synthesis approach makes it easy to covalently conjugate small peptides or fluorophores and enables control over crosslinking of different block co-polypeptides.

Self-assembly of diblock co-polypeptides containing dihydroxyphenylalanine to form polypeptide vesicles was demonstrated by Holowka and Deming.⁴¹ These vesicles, when chemically crosslinked, are stable under different conditions such as freeze-drying, organic solvents and osmotic stress. By using polar amino acids like lysine, glutamic acid, histidine and non-polar amino acids like phenylalanine, leucine, biomimetic vesicles with different physicochemical properties can be generated.^{42,43} Other hybrid vesicles called proteinosomes where proteins are covalently conjugated to polymers have been used to form biomimetic protocells⁴⁴ and prototissues.⁴⁵

2.2.2 Recombinantly generated polypeptides

Elastin-like polypeptides (ELPs) are amphiphilic polypeptides derived from tropoelastin that have found great utility in the fabrication of peptide-based cell membranes (discussed in section 4).⁴⁶ ELPs are comprised of alternate hydrophobic and hydrophilic domains that are made up of repeated units of pentapeptide Val-Pro-Gly-X-Gly, where X represents any canonical amino acid except proline. They undergo phase transition above a transition temperature (T_t) and aggregate into coacervates that are immiscible in water.⁴⁶ Below the T_t , they remain soluble in the aqueous phase. Exploiting this peculiar behavior, ELPs can be purified by inverse transition cycling method.⁴⁷

Repetitive nucleotide sequence encoding polypeptides like ELP, collagen-like protein, worm-silk-fibroin are difficult to synthesize. Conventional methods of gene synthesis fail to

construct the desired sequence due to the presence of repeated sequences. A number of solutions have been proposed to make viable genetic constructs and clone them into a suitable vector. These methods include concatemerization⁴⁸, overlap extension PCRs⁴⁹, one-vector-toolbox-platform⁵⁰, and recursive directional ligation (RDL)⁵¹, with RDL being the most popular method. In RDL, a small double-stranded DNA that serves as an ELP monomer is generated by ligating forward and reverse strands that are separately synthesized with their restriction site at respective ends, followed by insertion into a linearized vector with two different restriction sites (RE1 and RE2, see **Figure 2B**). ELP dimerization occurs in three steps; first, digestion of the vector with RE1 and RE2, which yields a ELP construct with complementary DNA overhangs; second, creation of a sticky-end with RE1 that already has an ELP monomer); and finally, ligation of ELP construct with sticky-end vector to yield a dimerized product. This method yields a low-efficiency ligation because of the circularization of insert. The overlap region required between the recognition sequence and the coding region limits the overall ligation efficiency.⁵² This ligation method was consequently improved and called recursive directional ligation by plasmid reconstitution (Pre-RDL).⁵³ Briefly, a parent plasmid is strategically cleaved into two parts by using different restriction sites. This results in two plasmids halves each containing an ELP construct. Then the two halves are ligated to generate a plasmid containing double the ELP gene. This method uses Type II restriction enzymes which cleave away from the recognition sequence, overcoming the limitation of RDL. ELPs have been extensively utilized in the field of drug-delivery⁵⁴ using different morphologies ranging from simple micelles⁵⁵ to nanofibers to beads-on-a-string.^{46,52,56–58,59}

3. Computational approaches to peptide self-assembly and design

3.1 All-atom and coarse-grained molecular simulation

Computer simulation has played an important role in the study of self-assembling peptides by providing microscopic insights into the molecular assembly mechanisms.⁶⁰ Simulations have been deployed to study the self-assembly of peptides over a wide range of length scales ranging from small dipeptides to long polypeptides. Appreciating that the size of the field precludes a comprehensive review, we highlight some selected studies of systems ordered from the small to the large. Self-assembly of dipeptides have been extensively studied with diphenylalanine (FF) peptide representing one of the smallest self-assembling peptides to have been investigated in simulations and experiments. Consistent with experimental findings,^{61–64} coarse-grained simulations of FF peptide performed by Guo et al. demonstrated the formation of a variety of nanostructures including vesicles, nanotubes and planar bilayers.⁶⁵

Interestingly, the *disassembly* of self-assembled FF crystals in buffer has been observed to help the formation of lipid vesicle in a dipeptide/phospholipid system.⁶⁶ Following this study, Guo et al. later found that triphenylalanine (FFF) peptide spontaneously forms solid nanospheres and nanotubes instead of water-filled vesicles and nanotubes formed by FF peptide.⁶⁷ They proposed that the interplay between side chain-side chain and main chain-main chain interactions is the key factor leading to the difference in the self-assembled structures between FFF and FF peptides.⁶⁷ Velichko et al. found using a coarse-grained model that the peptide amphiphile molecules self-assemble into a variety of structures, such as micelles and β -sheet, depending on the interplay between hydrophobic interactions and hydrogen bonding.⁶⁸ Lee et al. conducted all-atom simulation to examine the microscopic structure of cylindrical nanofiber formed by a peptide amphiphile whose peptide segment is SLSLAAAEIKVAV and determined the key energetic factors that contribute to the stability of fiber through the analysis of non-bonded interaction energy.⁶⁹ Fu et al. investigated the effects of temperature and hydrophobicity on the self-assembly of a peptide amphiphile, palmitoyl- $V_3A_3E_3$, and summarized the result in a phase-diagram⁷⁰ (**Figure 3A**). In a follow-up study, Fu et al. further investigated the kinetic pathway of palmitoyl- $V_3A_3E_3$ self-assembly under moderate temperature and different solvent conditions.⁷¹

In a series of studies, Shea and co-workers used coarse-grained simulations to investigate the effects of β -sheet propensity, surface interactions and pH on the fibrillization of peptides.^{72–75} They found that a decrease in β -sheet propensity leads to a decrease in fibril formation and an increase in the formation of toxic nonfibrillar structures.⁷⁴ They also proposed three mechanisms of fibril formation⁷² and showed that terminal charges play an important role in the aggregation morphology of a hydrophobic peptide YVIFL.⁷⁵ Lin et al. combined the experimental observation and computational simulation to confirm the micellization of PLGA-*b*-PPO-PLGA and PEG-*b*-PPO mixed chains where PLGA stands for poly(L-glutamic acid), PPO stands for poly(propylene oxide) and PEG stands for polyethylene glycol.⁷⁶ Using coarse-grained simulation, they were able to confirm that PPO blocks form the core of micelles while the PLGA and PEG blocks form the corona.⁷⁶

Although there have been many computational studies on the self-assembly of peptide copolymers in general, fewer computational studies have been conducted on the self-assembly of ELPs. Rauscher and Pomès performed long all-atom simulations for both the single (GVPGV)₇ chain and the nanoaggregate formed by multiple (GVPGV)₇ chains.⁷⁷ They found that single (GVPGV)₇ chain in solution samples disordered structures without strong preferred conformations and multiple (GVPGV)₇ chains assemble into nanoaggregates while showing

large degrees of conformational disorder.⁷⁷ Hassouneh et al. developed a theoretical model to describe the micellization of diblock ELP consisting of a hydrophobic block and a hydrophilic block.⁷⁸ They proposed a phase diagram showing various structures that can be formed by diblock ELP (**Figure 3B**), including weak micelles (micelles whose corona is almost unstretched), strong micelles (micelles whose corona is extended) and non-spherical morphologies. Their theoretical predictions of critical micelle temperature, hydrodynamic radius and aggregation number agree reasonably well with experiment.⁷⁸

Although not so many computational studies have been conducted directly on the self-assembly of ELPs, some studies have been conducted on the temperature-dependent conformational changes in ELPs. Tang et al. performed all-atom molecular dynamics simulations on a LG-ELP fused protein⁷⁹, where LG stands for laminin globular-like domain and the ELP sequence is (VPGKG)₂(VPGLG)₂(VPGIG)₂(VPGKG)₂. By performing the simulations over a range of physiological temperatures, it was found that the ELP region has a β -strand secondary structure between 310 K and 315 K.⁷⁹ Li et al. used all-atom simulations to investigate the mechanism of lower critical solution temperature (LCST) transition of (VPGVG)₁₈.⁸⁰ They found that this peptide shows β -turn structures over the temperature range 290–350 K and proposed that the LCST transition of (VPGVG)_n polypeptide results from an interplay between peptide-peptide and peptide-water interactions.⁸⁰ Zhao et al. also conducted all-atom simulations on (VPGVG)_n polypeptide and found that the number of hydrogen bonds between peptide and water as well as the number of water molecules in the first hydration shell of peptide backbone show clear transitions as temperature rises for single chain, suggesting that the LCST transition behavior is evident at single-molecule level.⁸¹ Tarakanova et al. conducted extensive molecular dynamics simulations to systematically investigate the effects of ionic concentration, chain length and sequence chemistry on the temperature-induced structural transitions in (VPGXG)_n polypeptide where X stands for an any amino acid except proline.⁸²

There have also been some studies on the effect of conjugation of ELP with collagen-like polypeptide (CLP) on the LCST-like transition temperature of ELP. Jayaraman and co-workers performed coarse-grained and all-atom simulations to probe the driving forces behind such effects for multiple ELP sequences.^{83,84} They proposed that conjugation of (VPGFG)₆ peptide with thermally-unresponsive stiff CLP triple helix leads to a decrease in the conformational entropy loss for polymer aggregation and thus lowers the transition temperature.⁸³ In a later study, they compared the transition temperatures of (VPGWG)₄ and (VPGFG)₄ peptides with and without conjugation to the same CLP sequence.⁸⁴ They found that the increased local stiffness and inter-molecular interaction of the W versus F substitution contribute to the

differences in the transition temperature of both free-state peptides and CLP-conjugated peptides.⁸⁴

3.2 Data-driven modeling and machine learning

In recent years, data-driven and machine learning (ML) techniques have come to play an increasingly important role in peptide property prediction and rational design. Kernel regression⁸⁵, support vector machine (SVM)⁸⁶ and artificial neural network⁸⁷, for example, have been deployed for the classification of peptides, prediction of peptide properties and design of novel peptide sequences that suit particular purposes. Leslie et al. proposed mismatch kernel based on a tree data structure to perform SVM classification of proteins for several benchmark tasks and demonstrated good performance.⁸⁸ Lee et al. used SVM to identify and discover new membrane-active and antimicrobial α -helical peptides.⁸⁹ Giguère et al. proposed a generic string kernel to perform kernel ridge regression to predict the peptide-protein binding affinity and demonstrated promising performance on several benchmark problems.⁸⁵ Thurston and Ferguson trained a quantitative structure-property relation model to perform extensive screening over π -conjugated oligopeptides and selected promising candidates that can self-assemble into nanoaggregates with desired optoelectronic properties.⁹⁰ Yang et al. employed doc2vec model⁹¹ from natural language processing to embed proteins into a vector space on which they performed Gaussian process regression (GPR) on benchmark tasks and nonlinear dimensionality reduction to evaluate the performance of protein embedding.⁹² Shmilovich et al. trained variational autoencoders to embed π -conjugated peptides into a low-dimensional vector space over which GPRs were constructed to support Bayesian optimization discovery of new π -conjugated peptides that can spontaneously form nanostructures with emergent optoelectronic properties.⁹³ In general, ML techniques can assist the discovery of novel promising materials by building predictive or generative models from existing data, thus making it a powerful tool to complement and guide simulation and experiment in the rational design of peptides that may be suitable as a novel chassis materials for synthetic cells.

4. Amphiphilic polypeptide vesicle as a compartment for cell-free expression

The recent demonstration of ELP vesicles and the emerging interest in cell-free expression (CFE) is fueling a new frontier in synthetic cell construction. CFE is a simple, rapid, and versatile tool for in situ protein synthesis with no requirement for purification.^{94–96} It is a powerful method for expressing any protein-of-interest outside the cell environment by harnessing the transcription-translation machinery of the cell.⁹⁷ This platform enables building or

conceptualizing a cell-like system from the bottom-up, including the incorporation of membrane proteins⁹⁸, reconstitution of cytoskeleton^{99,100}, and building artificial platelets.¹⁰¹ These previous studies were carried out using lipid bilayer vesicles. The enhanced robustness and functionality of peptide vesicles present new opportunities for CFE for operation in harsher environments or through a multiplexed functionality of the encapsulating membrane.

In a nutshell, a goal of synthetic cell research is to systematically build increasingly cell-like biomolecular systems that can sense a variety of external inputs such as small molecules, light or forces and respond by changing shape, synthesizing proteins, or altering internal enzymatic activities. In this context, it is crucial to have a robust compartment and ensure retention of the components inside a synthetic cell. Fatty acid and certain phospholipid vesicles generally have comparatively low retention efficiency of small molecule cargos than polymersomes and are often described as leaky because of their small membrane thickness.¹⁰² The high permeability of the lipid vesicles is due of the high lateral fluidity of the self-assembled low molecular weight (10^2 - 10^3 gmol⁻¹) lipid molecules. This can be problematic where cargo transport and confinement of small molecules are concerned. Since polymersomes are made of amphiphilic polymer of larger molecular weight on the order of 10^3 - 10^4 gmol⁻¹, they have relatively thick membrane size compared to liposomes, resulting in high retention efficiency and low permeability of cargos.^{103,104} The other important physical properties of membrane are bending rigidity and stretching elasticity. Lipid vesicle membrane typically has low bending rigidity of $\sim 20 k_B T$ in the fluid state, whereas polymersomes exhibit a range between 35–400 $k_B T$ depending on the membrane thickness. Stretching elasticity of membrane is associated with its lytic tension which is related to the toughness of the vesicles. Polymersomes are considered as tough vesicles because of their high lytic tension (20 - 30 mN m⁻¹) compared to that of liposomes (5 - 10 mN m⁻¹). For more details on the physical properties on liposomes and polymersomes, readers are referred to the review by Rideau et al.¹⁰⁵

Peptide vesicles fall under the amphiphilic block co-biopolymer category, and thus are expected to possess high retention efficiency, low permeability, and high lytic tension similar to polymersomes. To our knowledge, there have been no studies that have characterized physical properties for peptide vesicles. Interestingly, proteinosomes, a hybrid of protein and polymer (BSA-NH₂/PNIPAAm), reported by the Mann group and the de Greef group, showed permeability of polysaccharides (dextran up to 40 kDa)⁴⁴ and diffusion of unbound DNAs.¹⁰⁶ Their strategies of using protein/polymer hybrid presents an useful approach in fabrication of artificial tissue-like material and DNA-based communication in a population of synthetic cells.

As a first step of validation, peptide vesicles should be able to encapsulate and synthesize protein using a CFE reaction. Vogele et al. demonstrated the expression of protein in ELP vesicles encapsulating a bacterial transcription-translation (TX-TL) system.¹⁶ They showed the ELP vesicle size increases when the encapsulated ELP gene is expressed using a TX-TL system. ELP vesicles were prepared by thin film hydration method over glass beads followed by rehydration. The ELP is composed of hydrophilic domain (glutamic acid) and hydrophobic domain (phenylalanine) (see **Table 1**). This is a notable example of a peptide-based compartment exploited as a synthetic cell model. However, the size of vesicles is in the hundreds of nanometer range as depicted in **Figure 4A**. Another study, very recently reported by same group, demonstrated growth of ELP vesicles by fusion in an *in vitro* TX-TL reaction.¹⁰⁷ In this study, phenylalanine was used in the hydrophobic domain and arginine and glutamine were used in the hydrophilic domain (see **Figure 4B**). Using a THF swelling method where inner and outer solutions were introduced after removal of organic layer, the authors succeeded in forming micrometer size ELP vesicles.

5. Amphiphilic polypeptide in a protocell: An evolutionary remnant?

'How did life begin?'^{108,3} This question has led to many theories and speculations and it is still one of the most debated topics of all time. Life is speculated to have originated with the basic idea of a protocell – the self-assembly of molecules into a compartment with intrinsic genetic information and an ability to replicate. In the 1950's, the Miller-Urey experiments showed the formation of amino acids from essentially simple ingredients from early earth atmosphere like water, methane, ammonia and nitrogen.^{109,110} They found the synthesis of present cell membrane components such as fatty acids would be difficult from just mixing the simple gases in a reduced environment, supporting the notion that other forms of compartment likely existed before the advent of the lipid bilayer. For replication, it is postulated that protocells encapsulating RNAs could have started the non-enzymatic replication and catalysis which establishes the emergence of the RNA world.¹⁰ Given that the early earth environment supported synthesis of amino acids, is it conceivable that these may have assembled into prototypical peptide-based compartments that are more mechanically robust and relatively stable in extreme biochemical conditions compared to liposomes? The Schiller group has shown dynamic protein membrane formation using prebiotic amino acids that self-assemble (see **Figure 5A**).¹¹¹ They are basically modification of ELP with different guest residues for providing hydrophilic and hydrophobic properties (see **Table 1**). Solvent injection method using butanol/octanol was used to create organic-in-aqueous phase droplets in which the hydrophobic

part of ELP arranged on the interface and excess organic solvent was removed by dialysis.¹¹² Protein membrane compartments generated by this process can tolerate harsh conditions such as pH, temperature, and salt concentration. They demonstrated the cell-mimic processes with protein compartments like encapsulation of (bio)molecules, membrane fusion, and *in vitro* TX-TL (see **Figure 5A**). In this context, peptide vesicles combined with CFE systems may be more suitable for directed evolution experiments.

6. Challenges and opportunities

The emergence of peptide compartment as a synthetic cell chassis material presents tremendous opportunities but also poses new challenges. These materials were only demonstrated as a viable synthetic cell model within the past two years, meaning that detailed characterizations and compatibility with different compartmentalization approaches remained to be fully worked out. Peptide vesicles still require comprehensive testing and validation for encapsulation and cell-free expression efficiencies. It would also be interesting to explore the possibility of creating asymmetric bilayer vesicle consisting of a peptide and a lipid monolayer and investigate whether such membrane could allow the insertion of membrane proteins.

Although there are some distinct advantages of using a peptide compartment, there are some key challenges associated with conceptualizing a micron-sized peptide compartment. As their synthesis is concerned, the most effective and affordable method would be the recombinant approach where Pre-RDL can be used to generate desired polypeptides. Other methods are either difficult or not cost effective. For example, chemically synthesized peptides are relatively easy to control the side-chain modification but increasing amino acid length is a limitation due to synthesis cost. Researchers have fluorescently labelled polypeptides either by using unnatural amino acid or using a GFP-tag although the latter adds a significant size to the polypeptide chain.¹¹¹

Giant lipid vesicles and water-in-oil droplets have been utilized in directed evolution of proteins using *in vitro* protein synthesis^{113,114} including membrane proteins.^{115–117} The steps in a directed evolution experiment include encapsulation of DNA library, gene recovery by breaking emulsion or sorting droplets/liposomes by FACS, followed by re-encapsulation of DNA for the next cycle.¹¹⁸ Selection in directed evolution often involves vigorous mixing for washing and re-encapsulation steps, which always carry a risk of loss of molecules because of the fragile nature of lipid vesicles. Here, we propose ‘peptide vesicle display’ as a potential future direction for directed evolution study (see **Figure 5B**). Peptide vesicles can encapsulate different biomolecules such as protein, DNA, and cell-free expression components without any leakage through the membrane. These vesicles retain the cargo inside and are non-permeable to small

molecules like kanamycin as shown in **Figure 5A**, where kanamycin was introduced inside and/or outside the vesicle to inhibit the translation process. Although the porosity of peptide vesicles membrane is not well understood, it is expected to be less semi-permeable as compared to liposomes and polymersomes. Peptide vesicles are expected to be non-interfering with encapsulated entities and provide better handling at gene recovery step.

Recombinantly generated amphiphilic polypeptides usually self-assemble into micelles and do not form micron-sized vesicles on their own, unlike lipid vesicles. It requires a directed self-assembly to make peptide vesicle. The amphiphilic diblock, triblock or multiblock co-polypeptides could be a potential candidate for generation of the peptide compartment. Methods like gel-assisted GUV formation¹¹⁹ or microfluidics could be a potential alternative way of generating giant peptide vesicles.

Molecular simulation has proven to be a valuable tool in understanding and engineering self-assembling peptide vesicles^{65,67} and other structures.^{68,69,70,120,72–75,76} A body of simulation work has probed the conformational behavior of tropoelastin-derived ELPs with a particular focus on the LCST transition.^{79,81,80,82,84} In future work, we anticipate that coarse-grained molecular models of ELPs^{83,84,121} combined with enhanced sampling and free energy calculations^{122–130} present a means to directly simulate the stability and assembly of ELP peptide compartments. The high computational cost and vast sequence space of candidate ELPs make integrating these calculations with data-driven Bayesian optimization and active learning techniques particularly valuable in rationally traversing sequence space and focusing computational resources on the simulation of self-assembling ELPs with the most desirable emergent thermodynamic or structural properties.^{93,131–138}

In sum, the demonstrated fabrication of peptide-based ELP vesicles with a CFE system pioneered by Voegelé et al.¹⁶ and Schreiber et al.¹¹¹ have opened up tremendous opportunities for exploring biomolecular reactions in cell-sized peptide compartments. Such highly stable peptide compartments could provide a platform for origin of life study and highly tunable, multifunctional, multi-compartment systems in the field of bottom-up synthetic biology.

Table of acronyms

Acronym	Meaning
FF	Diphenylalanine
FFF	Triphenylalanine
LCST	Lower critical solution temperature
CLP	Collagen-like polypeptide
ELP	Elastin-like polypeptide
ML	Machine Learning
SVM	Support vector machine
GPR	Gaussian process regression
RDL	Recursive directional ligation
Pre-RDL	Recursive directional ligation by plasmid reconstitution
CFE	Cell-free expression
TX-TL	Transcription-translation
PNIPAAm	Poly(N-isopropylacrylamide)
THF	Tetrahydrofuran
IVTT	<i>In vitro</i> transcription-translation
Kan	Kanamycin
PMBC	Protein membrane-based compartments

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant Nos. DMR-1939534 (APL) and DMR-1939463 (ALF). The authors thank Yashar Bashirzadeh for valuable comments and discussion.

Figure captions

Figure 1: Naturally occurring amphiphilic proteins. (A) i) Schematic representation of giant vesicle formation from organic-in-water droplets. ii) Recombinant expression of oleosin mutant in *E. coli* and self-assembly of oleosin upon solvent injection. All the mutants have fixed hydrophobic regions while hydrophilic ends were modified. Brightfield and fluorescence images with red as Nile Red and green as calcein. Scale bars are 50 and 5 μm in the upper and bottom panels, respectively. Figures in ii) are reproduced from ref. 26 with permission from National Academy of Sciences, copyright 2012. (B) Self-assembly of hydrophobins in water-in-oil droplets and oil-in-water droplets using a microfluidic device. Reproduced from ref. 30 with permission from John Wiley and Sons, copyright 2016.

Figure 2: Different polypeptide synthesis platforms. (A) General chemical method of synthesis of polypeptides for formation of vesicles (like polymersomes). (B) i) Schematic of recursive directional ligation. It is a method to increase the desired size of a protein (especially ELP) by doubling the length of an oligomer gene using same vector. Reproduced from ref.53 with permission from American Chemical Society, copyright 2010. ii) Other method of making proteins with repetitive sequence are concatemerization (ligating different fragments of oligomer gene) and overlap extension PCRs (two ssDNAs are annealed to each other through overlap region and extended during PCRs).

Figure 3: Molecular simulation in peptide self-assembly. (A) Computational phase diagram of the equilibrium structures self-assembled from palmitoyl- $V_3A_3E_3$. A variety of structures can be formed including micelles, nanofiber, oligomers, and random coils. Reproduced from ref.70 with permission from American Chemical Society, copyright 2014. (B) Phase diagram of diblock ELP predicted by *Hassounah et al.* N_A and N_B are the degrees of polymerization for hydrophilic and hydrophobic blocks, respectively. The phase diagram includes monomer (region I), weak micelle (region II), strong micelle (region III) and non-spherical aggregate (region IV). Reproduced from ref.78 with permission from American Chemical Society, copyright 2015.

Figure 4: ELP vesicle as an artificial cell model. (A) i) Glass bead-mediated vesicle formation by rehydrating dried ELP. The ELP here is a diblock amphiphilic polypeptide with glutamic acid as hydrophilic domain and phenylalanine as hydrophobic domain. ii) *left*, dynamic light scattering and transmission electron microscopy results showed the average size of ELP vesicles of 176 nm, and cell-free protein expression of mVenus inside an ELP vesicle (*right*). Reproduced from ref.16 with permission from Springer Nature, copyright 2018. (B) i) Sequence design of amphiphilic ELP, using arginine and glutamine as hydrophilic domain and phenylalanine as hydrophobic domain. ELP giant vesicles generated by THF swelling method. ii) Schematic illustration of growth of ELP vesicles by fusion of vesicles containing components of transcription mix that eventually started the synthesis of dBroc coli RNA aptamers. Reproduced from ref.107 published under a Creative Commons license.

Figure 3: ELP vesicles as a protocell model. (A) i) Epifluorescence and TEM images of self-assembly of protein membrane-based compartments (PMBCs) from diblock amphiphilic ELP-

derivative His-mEGFP-H40I30 (*left*), His-mEGFP-S40I30 (*right*). mEGFP tag was used for visualization. ii) Red fluorescence image (*middle*) showing encapsulation of mCherry in PMBCs with its proper folding during self-assembly. iii) *in vitro* transcription-translation (IVTT) and membrane incorporation of mCherry-H40I30 membrane block in an assembled PMBCs from His-mEGFP-H40I30. Kanamycin (Kan) was used as IVTT inhibitor. Red fluorescence shows the IVTT of mCherry-H40I30 inside and outside (agglomerates seen) of the PMBCs in absence of Kan (*left panel*). In the middle panel, where Kan was added to the outer solution of PMBCs inhibiting IVTT, led to disappearance in red fluorescence agglomerates. No fluorescence observed in the right panel because of the inhibition of IVTT due to the presence of Kan inside and outside of PMBCs. iv) Fusion of BDP-K40I30 PMBCs (3 and 14 μm) recorded at 4 *fps*. Scale bars are 5 μm and 100 nm for all the epifluorescence and TEM images, respectively. Reproduced from ref.111 with permission from American Chemical Society, copyright 2019. (B) Schematic overview of the proposed concept of 'peptide vesicle display' showing preparation of gene library and its encapsulation inside peptide vesicle with cell-free expression. Sorting of desired fluorescence vesicles using FACS, followed by gene recovery and preparation of the library for next round.

References:

- 1 C. Xu, S. Hu and X. Chen, *Mater. Today*, 2016, **19**, 516–532.
- 2 C. Chiarabelli, P. Stano and P. L. Luisi, *Curr. Opin. Biotechnol.*, 2009, **20**, 492–497.
- 3 J. P. Schrum, T. F. Zhu and J. W. Szostak, *Cold Spring Harb. Perspect. Biol.*, 2010, **2**, a002212–a002212.
- 4 E. Godino, J. N. López, D. Foschepoth, C. Cleij, A. Doerr, C. F. Castellà and C. Danelon, *Nat. Commun.*, 2019, **10**, 1–12.
- 5 N. Y. Kostina, K. Rahimi, Q. Xiao, T. Haraszti, S. Dedisch, J. P. Spatz, U. Schwaneberg, M. L. Klein, V. Percec, M. Möller and C. Rodriguez-Emmenegger, *Nano Lett.*, 2019, **19**, 5732–5738.
- 6 D. Konetski, D. Zhang, D. K. Schwartz and C. N. Bowman, *Chem. Mater.*, 2018, **30**, 8757–8763.
- 7 A. S. Cans, N. Wittenberg, R. Karlsson, L. Sombers, M. Karlsson, O. Orwar and A. Ewing, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 400–404.
- 8 Y. Elani, T. Trantidou, D. Wylie, L. Dekker, K. Polizzi, R. V. Law and O. Ces, *Sci. Rep.*, 2018, **8**, 4564.
- 9 Y. C. Tan, K. Hettiarachchi, M. Siu, Y. R. Pan and A. P. Lee, *J. Am. Chem. Soc.*, 2006, **128**, 5656–5658.
- 10 J. C. Blain and J. W. Szostak, *Annu. Rev. Biochem.*, 2014, **83**, 615–640.
- 11 T. M. Allen and P. R. Cullis, *Science (80-)*, 2004, **303**, 1818–1822.
- 12 N. Tandel, A. Z. Joseph, A. Joshi, P. Shrama, R. P. Mishra, R. K. Tyagi and P. S. Bisen, *Expert Rev. Mol. Diagn.*, 2020, **20**, 533–541.
- 13 B. Iyisan and K. Landfester, *Macromol. Rapid Commun.*, 2019, **40**, 1800577.
- 14 B. M. Discher, Y. Y. Won, D. S. Ege, J. C. M. Lee, F. S. Bates, D. E. Discher and D. A. Hammer, *Science (80-)*, 1999, **284**, 1143–1146.
- 15 M. L. Jacobs, M. A. Boyd and N. P. Kamat, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 4031–4036.
- 16 K. Vogeles, T. Frank, L. Gasser, M. A. Goetzfried, M. W. Hackl, S. A. Sieber, F. C. Simmel and T. Pirzer, *Nat. Commun.*, 2018, **9**, 1–7.
- 17 A. Cooper and M. W. Kennedy, *Biophys. Chem.*, 2010, **151**, 96–104.
- 18 Q. Ren, A. H. Kwan and M. Sunde, *Biopolymers*, 2013, **100**, 601–612.
- 19 M. Schor, J. L. Reid, C. E. MacPhee and N. R. Stanley-Wall, *Trends Biochem. Sci.*, 2016, **41**, 610–620.

- 20 Q. Shao, X. Liu, T. Su, C. Ma and P. Wang, *Front. Plant Sci.*, 2019, 10, 1568.
- 21 A. H. C. Huang, *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, 1992, **43**, 177–200.
- 22 O. Leprince, A. C. Van Aelst, H. W. Pritchard and D. J. Murphy, *Planta*, 1997, **204**, 109–119.
- 23 D. J. Lacey, N. Wellner, F. Beaudoin, J. A. Napier and P. R. Shewry, *Biochem. J.*, 1998, **334**, 469–477.
- 24 K. Hsieh and A. H. C. Huang, *Plant Physiol.*, 2004, 136, 3427–3434.
- 25 F. Beaudoin and J. A. Napier, *Planta*, 2002, **215**, 293–303.
- 26 K. B. Vargo, R. Parthasarathy and D. A. Hammer, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 11657–11662.
- 27 M. B. Linder, G. R. Szilvay, T. Nakari-Setälä and M. E. Penttilä, *FEMS Microbiol. Rev.*, 2005, **29**, 877–896.
- 28 H. A. B. Wösten and J. G. H. Wessels, *Mycoscience*, 1997, **38**, 363–374.
- 29 V. Lo, J. I-Chun Lai and M. Sunde, *Adv. Exp. Med. Biol.*, 2019, **1174**, 161–185.
- 30 H. Hähl, J. N. Vargas, A. Griffo, P. Laaksonen, G. Szilvay, M. Lienemann, K. Jacobs, R. Seemann and J. B. Fleury, *Adv. Mater.*, 2017, **29**, 1602888.
- 31 A. Cooper, M. W. Kennedy, R. I. Fleming, E. H. Wilson, H. Videler, D. L. Wokosin, T. J. Su, R. J. Green and J. R. Lu, *Biophys. J.*, 2005, **88**, 2114–2125.
- 32 D. Cavalcante Hissa, G. Arruda Bezerra, R. Birner-Gruenberger, L. Paulino Silva, I. Usön, K. Gruber and V. M. Maclel Melo, *ChemBioChem*, 2014, **15**, 393–398.
- 33 C. D. Mackenzie, B. O. Smith, A. Meister, A. Blume, X. Zhao, J. R. Lu, M. W. Kennedy and A. Cooper, *Biophys. J.*, 2009, **96**, 4984–4992.
- 34 D. Wendell, J. Todd and C. Montemagno, *Nano Lett.*, 2010, **10**, 3231–3236.
- 35 J. G. Beeley, R. Eason and D. H. Snow, *Biochem. J.*, 1986, **235**, 645–650.
- 36 R. E. McDonald, R. I. Fleming, J. G. Beeley, D. L. Bovell, J. R. Lu, X. Zhao, A. Cooper and M. W. Kennedy, *PLoS One*, 2009, **4**, 1–12.
- 37 M. W. Kennedy, *Biochem. Soc. Trans.*, 2011, **39**, 1017–1022.
- 38 S. J. Vance, R. E. McDonald, A. Cooper, B. O. Smith and M. W. Kennedy, *J. R. Soc. Interface*, 2013, **10**, 20130453.
- 39 V. Breedveld, A. P. Nowak, J. Sato, T. J. Deming and D. J. Pine, *Macromolecules*, 2004, **37**, 3943–3953.
- 40 M. Yu, A. P. Nowak, T. J. Deming and D. J. Pochan, *J. Am. Chem. Soc.*, 1999, **121**, 12210–12211.
- 41 E. P. Holowka and T. J. Deming, *Macromol. Biosci.*, 2010, **10**, 496–502.

- 42 E. P. Holowka, D. J. Pochan and T. J. Deming, *J. Am. Chem. Soc.*, 2005, **127**, 12423–12428.
- 43 A. Sulistio, A. Blencowe, J. Wang, G. Bryant, X. Zhang and G. G. Qiao, *Macromol. Biosci.*, 2012, **12**, 1220–1231.
- 44 X. Huang, M. Li, D. C. Green, D. S. Williams, A. J. Patil and S. Mann, *Nat. Commun.*, 2013, **4**, 1–9.
- 45 P. Gobbo, A. J. Patil, M. Li, R. Harniman, W. H. Briscoe and S. Mann, *Nat. Mater.*, 2018, **17**, 1145–1153.
- 46 D. H. T. Le and A. Sugawara-Narutaki, *Mol. Syst. Des. Eng.*, 2019, **4**, 545–565.
- 47 D. E. Meyer and A. Chilkoti, *Nat. Biotechnol.*, 1999, **17**, 1112–1115.
- 48 M. J. White, B. W. Fristensky and W. F. Thompson, *Anal. Biochem.*, 1991, **199**, 184–190.
- 49 M. Amiram, F. G. Quiroz, D. J. Callahan and A. Chilkoti, *Nat. Mater.*, 2011, **10**, 141–148.
- 50 M. C. Huber, A. Schreiber, W. Wild, K. Benz and S. M. Schiller, *Biomaterials*, 2014, **35**, 8767–8779.
- 51 D. E. Meyer and A. Chilkoti, *Biomacromolecules*, 2002, **3**, 357–367.
- 52 S. Saha, S. Banskota, S. Roberts, N. Kirmani and A. Chilkoti, *Adv. Ther.*, 2020, **3**, 1900164.
- 53 J. R. McDaniel, J. A. MacKay, F. G. Quiroz and A. Chilkoti, *Biomacromolecules*, 2010, **11**, 944–952.
- 54 J. C. Rodríguez-Cabello, F. J. Arias, M. A. Rodrigo and A. Girotti, *Adv. Drug Deliv. Rev.*, 2016, **97**, 85–100.
- 55 W. M. Park and J. A. Champion, *J. Am. Chem. Soc.*, 2014, **136**, 17906–17909.
- 56 A. K. Varanko, J. C. Su and A. Chilkoti, *Annu. Rev. Biomed. Eng.*, 2020, **22**, 343–369.
- 57 S. R. MacEwan, I. Weitzhandler, I. Hoffmann, J. Genzer, M. Gradzielski and A. Chilkoti, *Biomacromolecules*, 2017, **18**, 599–609.
- 58 S. Roberts, V. Miao, S. Costa, J. Simon, G. Kelly, T. Shah, S. Zauscher and A. Chilkoti, *Nat. Commun.*, 2020, **11**, 1342.
- 59 D. G. Fatouros, D. A. Lamprou, A. J. Urquhart, S. N. Yannopoulos, I. S. Vizirianakis, S. Zhang and S. Koutsopoulos, *ACS Appl. Mater. Interfaces*, 2014, **6**, 8184–8189.
- 60 C. Yuan, S. Li, Q. Zou, Y. Ren and X. Yan, *Phys. Chem. Chem. Phys.*, 2017, **19**, 23614–23631.
- 61 M. Reches and E. Gazit, *Science (80-.)*, 2003, **300**, 625–627.
- 62 X. Yan, P. Zhu and J. Li, *Chem. Soc. Rev.*, 2010, **39**, 1877–1890.
- 63 C. H. Görbitz, *Chem. Commun.*, 2006, 2332–2334.

- 64 M. Reches and E. Gazit, *Nano Lett.*, 2004, **4**, 581–585.
- 65 C. Guo, Y. Luo, R. Zhou and G. Wei, *ACS Nano*, 2012, **6**, 3907–3918.
- 66 M. Fu and J. Li, *Angew. Chemie Int. Ed.*, 2018, **57**, 11404–11407.
- 67 C. Guo, Y. Luo, R. Zhou and G. Wei, *Nanoscale*, 2014, **6**, 2800–2811.
- 68 Y. S. Velichko, S. I. Stupp and M. O. de la Cruz, *J. Phys. Chem. B*, 2008, **112**, 2326–2334.
- 69 O. S. Lee, S. I. Stupp and G. C. Schatz, *J. Am. Chem. Soc.*, 2011, **133**, 3677–3683.
- 70 I. W. Fu, C. B. Markegard, B. K. Chu and H. D. Nguyen, *Langmuir*, 2014, **30**, 7745–7754.
- 71 I. W. Fu, C. B. Markegard and H. D. Nguyen, *Langmuir*, 2015, **31**, 315–324.
- 72 G. Bellesia and J. E. Shea, *J. Chem. Phys.*, 2009, **131**, 111102.
- 73 A. Morriss-Andrews, G. Bellesia and J.-E. Shea, *J. Chem. Phys.*, 2011, **135**, 85102.
- 74 G. Bellesia and J.-E. Shea, *J. Chem. Phys.*, 2009, **130**, 145103.
- 75 T. D. Do, N. E. LaPointe, N. J. Economou, S. K. Buratto, S. C. Feinstein, J.-E. Shea and M. T. Bowers, *J. Phys. Chem. B*, 2013, **117**, 10759–10768.
- 76 J. Lin, J. Zhu, T. Chen, S. Lin, C. Cai, L. Zhang, Y. Zhuang and X.-S. Wang, *Biomaterials*, 2009, **30**, 108–117.
- 77 S. Rauscher and R. Pomès, *Elife*, 2017, **6**, e26526.
- 78 W. Hassouneh, E. B. Zhulina, A. Chilkoti and M. Rubinstein, *Macromolecules*, 2015, **48**, 4183–4195.
- 79 J. D. Tang, C. E. McAnany, C. Mura and K. J. Lampe, *Biomacromolecules*, 2016, **17**, 3222–3233.
- 80 N. K. Li, F. G. Quiroz, C. K. Hall, A. Chilkoti and Y. G. Yingling, *Biomacromolecules*, 2014, **15**, 3522–3530.
- 81 B. Zhao, N. K. Li, Y. G. Yingling and C. K. Hall, *Biomacromolecules*, 2016, **17**, 111–118.
- 82 A. Tarakanova, W. Huang, A. S. Weiss, D. L. Kaplan and M. J. Buehler, *Biomaterials*, 2017, **127**, 49–60.
- 83 J. E. Condon, T. B. Martin and A. Jayaraman, *Soft Matter*, 2017, **13**, 2907–2918.
- 84 A. Prhashanna, P. A. Taylor, J. Qin, K. L. Kiick and A. Jayaraman, *Biomacromolecules*, 2019, **20**, 1178–1189.
- 85 S. Giguère, M. Marchand, F. Laviolette, A. Drouin and J. Corbeil, *BMC Bioinformatics*, 2013, **14**, 82.
- 86 E. Y. Lee, G. C. L. Wong and A. L. Ferguson, *Bioorganic Med. Chem.*, 2018, **26**, 2708–2718.
- 87 M. Nielsen and O. Lund, *BMC Bioinformatics*, 2009, **10**, 296.

- 88 C. S. Leslie, E. Eskin, A. Cohen, J. Weston and W. S. Noble, *Bioinformatics*, 2004, **20**, 467–476.
- 89 E. Y. Lee, B. M. Fulan, G. C. L. Wong and A. L. Ferguson, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 13588–13593.
- 90 B. A. Thurston and A. L. Ferguson, *Mol. Simul.*, 2018, **44**, 930–945.
- 91 Q. Le and T. Mikolov, in *31st International Conference on Machine Learning, ICML 2014*, eds. E. P. Xing and T. Jebara, PMLR, Beijing, China, 2014, vol. 4, pp. 2931–2939.
- 92 K. K. Yang, Z. Wu, C. N. Bedbrook and F. H. Arnold, *Bioinformatics*, 2018, **34**, 2642–2648.
- 93 K. Shmilovich, R. A. Mansbach, H. Sidky, O. E. Dunne, S. S. Panda, J. D. Tovar and A. L. Ferguson, *J. Phys. Chem. B*, 2020, **124**, 3873–3891.
- 94 K. Khambhati, G. Bhattacharjee, N. Gohil, D. Braddick, V. Kulkarni and V. Singh, *Front. Bioeng. Biotechnol.*, 2019, **7**, 248.
- 95 N. E. Gregorio, M. Z. Levine and J. P. Oza, *Methods Protoc.*, 2019, **2**, 24.
- 96 V. Noireaux and A. P. Liu, *Annu. Rev. Biomed. Eng.*, 2020, **22**, 51–77.
- 97 Z. Z. Sun, C. A. Hayes, J. Shin, F. Caschera, R. M. Murray and V. Noireaux, *J. Vis. Exp.*, 2013, 50762.
- 98 S. Majumder, J. Garamella, Y. L. Wang, M. Denies, V. Noireaux and A. P. Liu, *Chem. Commun.*, 2017, **53**, 7349–7352.
- 99 J. Garamella, S. Majumder, A. P. Liu and V. Noireaux, *ACS Synth. Biol.*, 2019, **8**, 1913–1920.
- 100 Y. T. Maeda, T. Nakadai, J. Shin, K. Uryu, V. Noireaux and A. Libchaber, *ACS Synth. Biol.*, 2012, **1**, 53–59.
- 101 S. Majumder and A. P. Liu, *Phys. Biol.*, 2017, **15**, 013001.
- 102 H. Che and J. C. M. Van Hest, *J. Mater. Chem. B*, 2016, **4**, 4632–4647.
- 103 D. E. Discher and F. Ahmed, *Annu. Rev. Biomed. Eng.*, 2006, **8**, 323–341.
- 104 J. F. Le Meins, O. Sandre and S. Lecommandoux, *Eur. Phys. J. E*, 2011, **34**, 1–17.
- 105 E. Rideau, R. Dimova, P. Schwille, F. R. Wurm and K. Landfester, *Chem. Soc. Rev.*, 2018, **47**, 8572–8610.
- 106 A. Joesaar, S. Yang, B. Bögels, A. van der Linden, P. Pieters, B. V. V. S. P. Kumar, N. Dalchau, A. Phillips, S. Mann and T. F. A. de Greef, *Nat. Nanotechnol.*, 2019, **14**, 369–378.
- 107 T. Frank, K. Vogele, A. Dupin, F. C. Simmel and T. Pirzer, *Chem. – A Eur. J.*, 2020, **13**, chem.202003366.

- 108 J. Szostak, *Nature*, 2018, **557**, S13–S15.
- 109 H. C. Urey, *Proc. Natl. Acad. Sci.*, 1952, **38**, 351–363.
- 110 S. L. Miller, *Science (80-)*, 1953, **117**, 528–529.
- 111 A. Schreiber, M. C. Huber and S. M. Schiller, *Langmuir*, 2019, **35**, 9593–9610.
- 112 A. Schreiber, L. G. Stühn, S. E. Geissinger, M. C. Huber and S. M. Schiller, *J. Vis. Exp.*, 2020, **2020**, e60935.
- 113 T. Nishikawa, T. Sunami, T. Matsuura and T. Yomo, *J. Nucleic Acids*, 2012, 2012, 11.
- 114 H. Sugiyama and T. Toyota, *Life*, 2018, **8**, 53.
- 115 S. Fujii, T. Matsuura, T. Sunami, Y. Kazuta and T. Yomo, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 16796–16801.
- 116 S. Fujii, T. Matsuura, T. Sunami, T. Nishikawa, Y. Kazuta and T. Yomo, *Nat. Protoc.*, 2014, **9**, 1578–1591.
- 117 A. Uyeda, S. Nakayama, Y. Kato, H. Watanabe and T. Matsuura, *Anal. Chem.*, 2016, **88**, 12028–12035.
- 118 D. S. Tawfik and A. D. Griffiths, *Nat. Biotechnol.*, 1998, **16**, 652–656.
- 119 A. C. Greene, D. Y. Sasaki and G. D. Bachand, *J. Vis. Exp.*, 2016, **2016**, 54051.
- 120 I. W. Fu and H. D. Nguyen, *Biomacromolecules*, 2015, **16**, 2209–2219.
- 121 S. J. Marrink and D. P. Tieleman, *Chem. Soc. Rev.*, 2013, **42**, 6801–6822.
- 122 C. Abrams and G. Bussi, *Entropy*, 2014, **16**, 163–199.
- 123 Y. Miao and J. A. McCammon, *Mol. Simul.*, 2016, **42**, 1046–1055.
- 124 Y. I. Yang, Q. Shao, J. Zhang, L. Yang and Y. Q. Gao, *J. Chem. Phys.*, 2019, **151**, 070902.
- 125 O. Valsson and M. Parrinello, *Phys. Rev. Lett.*, 2014, **113**, 1–5.
- 126 H. Sidky, W. Chen and A. L. Ferguson, *Mol. Phys.*, 2020, **118**, e1737742.
- 127 A. Pohorille, C. Jarzynski and C. Chipot, *J. Phys. Chem. B*, 2010, **114**, 10235–10253.
- 128 M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia and M. Parrinello, *Comput. Phys. Commun.*, 2009, **180**, 1961–1972.
- 129 H. Sidky, Y. J. Colón, J. Helfferich, B. J. Sikora, C. Bezik, W. Chu, F. Giberti, A. Z. Guo, X. Jiang, J. Lequieu, J. Li, J. Moller, M. J. Quevillon, M. Rahimi, H. Ramezani-Dakhel, V. S. Rathee, D. R. Reid, E. Sevgen, V. Thapar, M. A. Webb, J. K. Whitmer and J. J. De Pablo, *J. Chem. Phys.*, 2018, **148**, 044104.
- 130 C. Chipot and A. Pohorille, *Free Energy Calculations*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2001, vol. 86.

- 131 E. Brochu, V. M. Cora and N. de Freitas, .
- 132 E. P. Fox and D. S. Sivia, *Technometrics*, 1998, **40**, 155.
- 133 C. Kim, A. Chandrasekaran, A. Jha and R. Ramprasad, *MRS Commun.*, 2019, **9**, 860–866.
- 134 J. Ling, M. Hutchinson, E. Antono, S. Paradiso and B. Meredig, *Integr. Mater. Manuf. Innov.*, 2017, **6**, 207–217.
- 135 R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams and A. Aspuru-Guzik, *ACS Cent. Sci.*, 2018, **4**, 268–276.
- 136 G. R. Bickerton, G. V. Paolini, J. Besnard, S. Muresan and A. L. Hopkins, *Nat. Chem.*, 2012, **4**, 90–98.
- 137 D. Xue, P. V. Balachandran, J. Hogden, J. Theiler, D. Xue and T. Lookman, *Nat. Commun.*, 2016, **7**, 1–9.
- 138 R. Yuan, Z. Liu, P. V. Balachandran, D. Xue, Y. Zhou, X. Ding, J. Sun, D. Xue and T. Lookman, *Adv. Mater.*, 2018, **30**, 1–8.

Figure 1

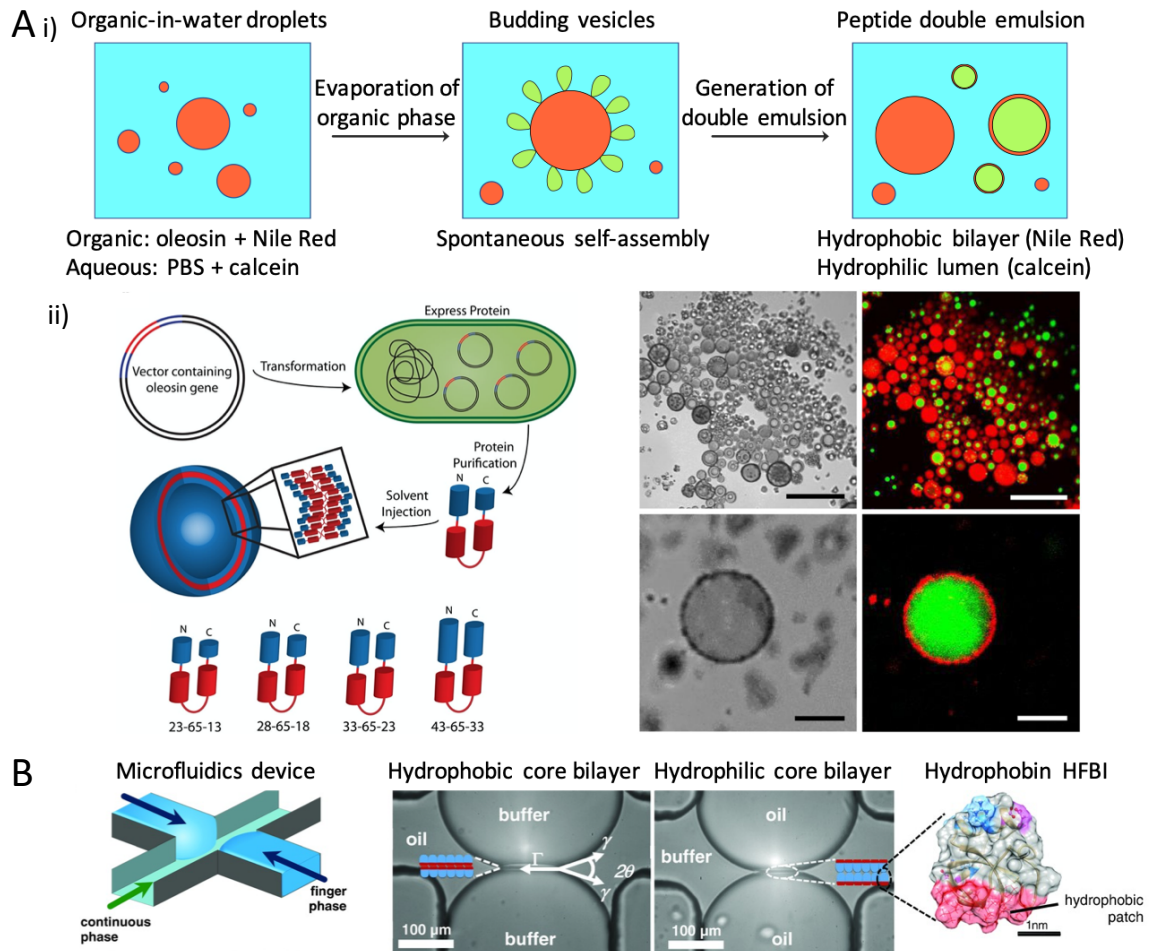


Figure 2

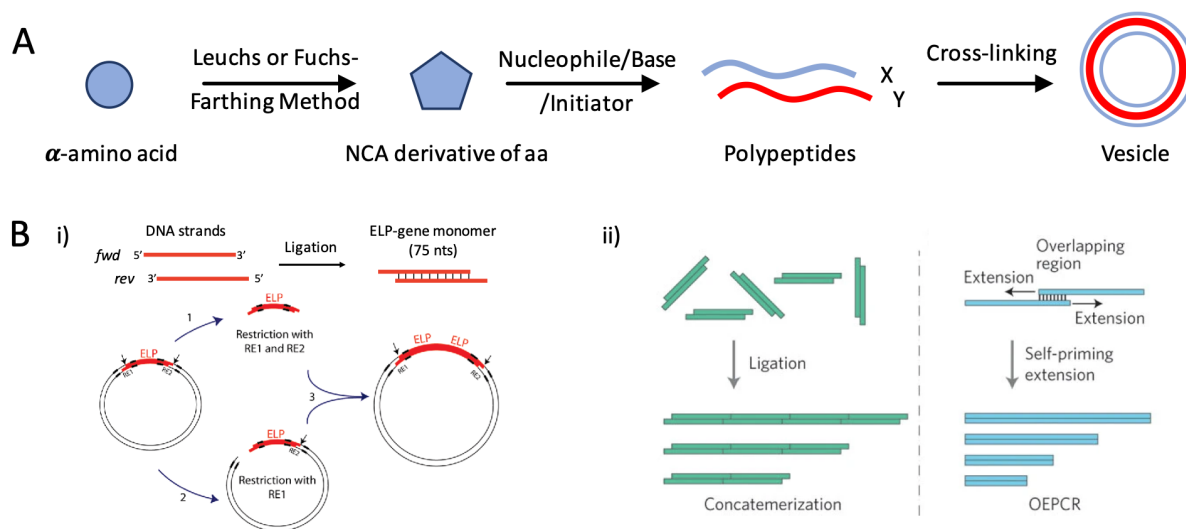


Figure 3

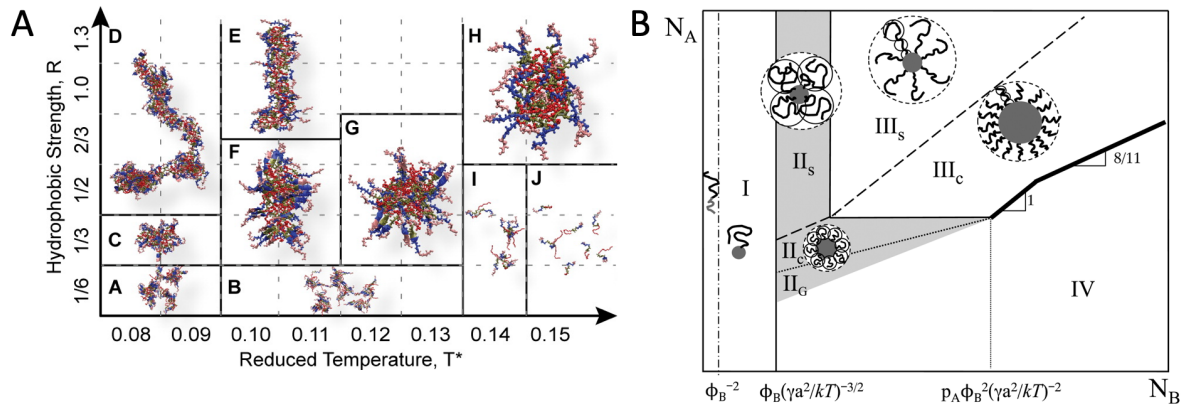


Figure 4

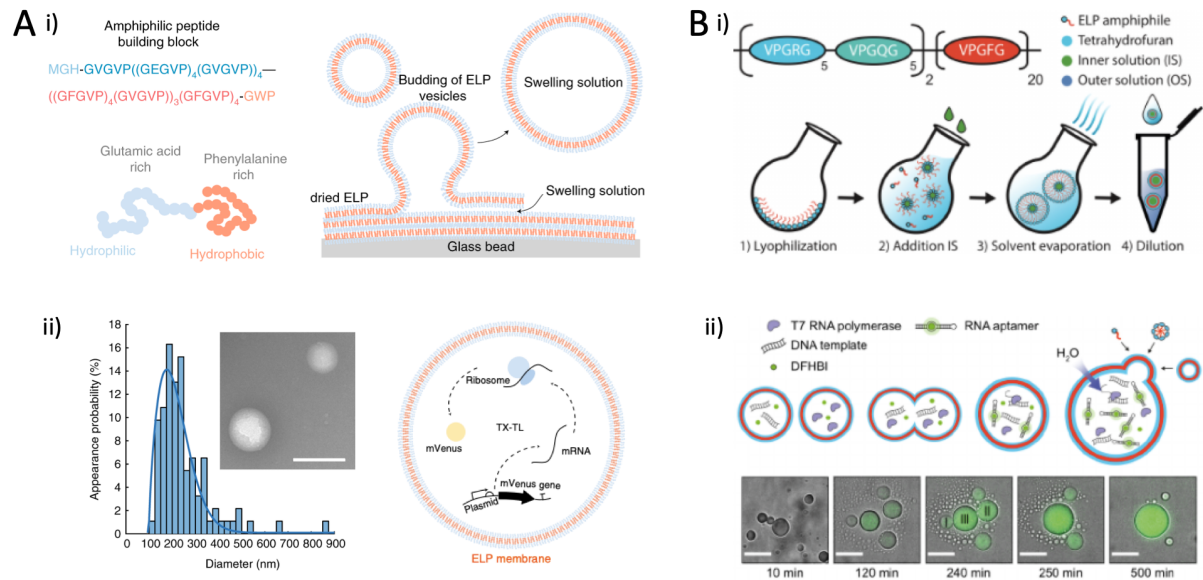


Figure 5

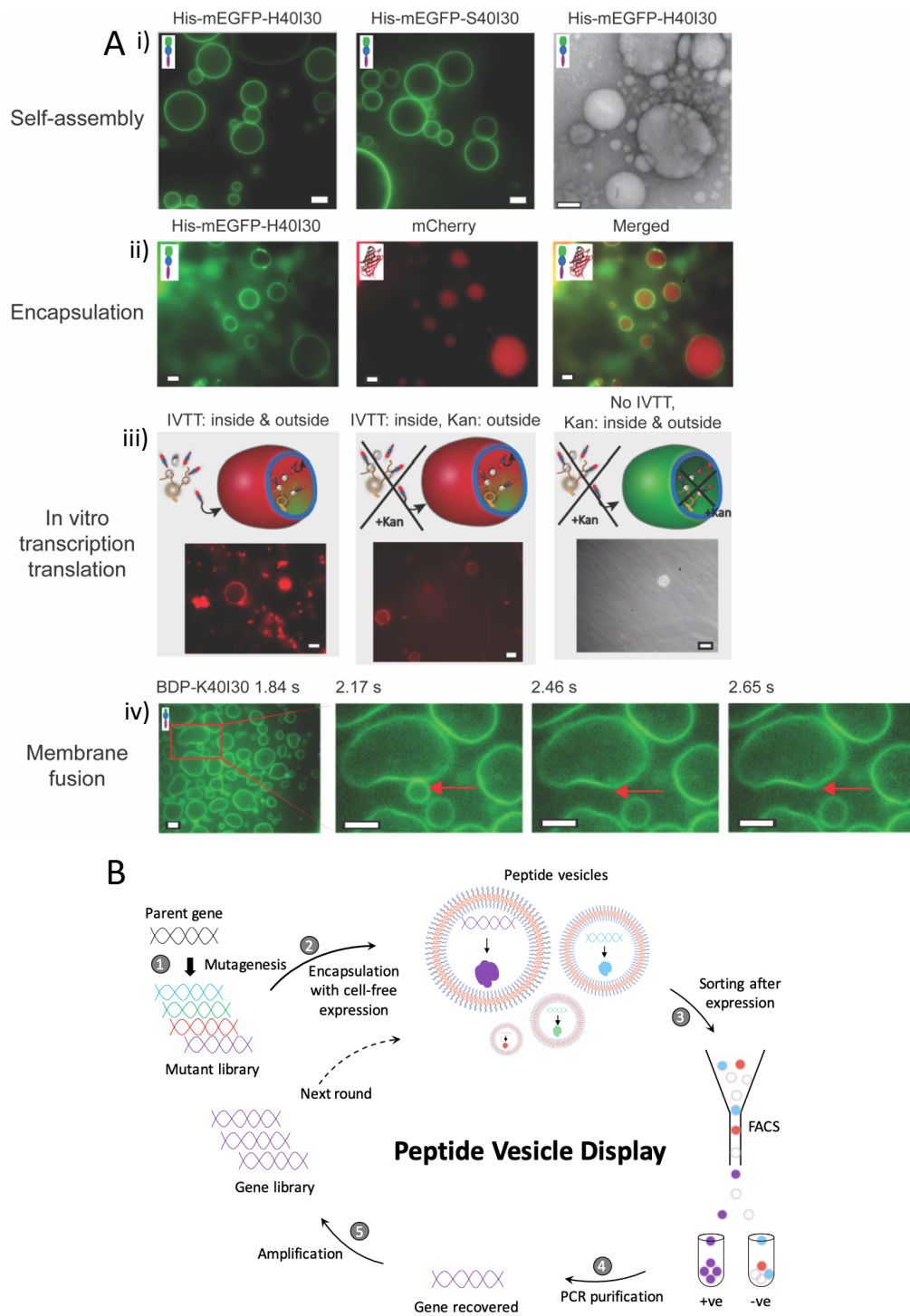


Table 1: List of self-assembling amphiphilic polypeptides/proteins

Self-assembled polypeptides	Synthesis/ origin	Sequence	Guest residues	Amino acid residue (total)	Size	Application	Ref
Oleosins	Natural. Plant oil bodies (Safflower)	Wildtype: N-P-C (42-87-63)	(Hydrophilic)- (hydrophobic)- (hydrophilic)*	192	nm - μm	Drug vehicles	26
Hydrophobins	Natural. Globular proteins from filamentous fungi	HFBI	Amphiphilic*	100	30-60 μm ; (Oil)	Drug vehicles, food industry, medical implant biocompatibility	30
L-DOPA	Chemical	L-3,4-DOPA residues, i.e. $\text{K}_{60}(\text{DOPA}_x/\text{L}_y)_{20}$	Block co- polypeptide amphiphile	$\text{K}_{60}(\text{DOPA}_{0.25}/\text{L}_{0.75})_{20}$ & $\text{K}_{60}\text{DOPA}_{20}$	3-5 μm	High stability towards osmotic-shock, organic solvent, media-serum proteins	41
ELP oligoarginine	Recombinant	Natural IDP- tropoelastin or partially ordered polymers (POPs)	Valine (V)	POP(V)-25% ELP (V_4A_1) [#]	40-50 μm	Unique micro- architecture with biocompatible supramolecular aggregates	58
ELP-EF	Recombinant	Peptidosome derivative of elastin-like peptide (ELP)	Diblock copolymer. Glutamic acid (E) & Phenylalanine (F)	$\text{E}_{16}\text{F}_{16}$;	175-200 nm	Cell-free expression of ELP and assembly in vesicle	16
$(\text{R}_5\text{Q}_5)_2\text{F}_{20}$	Recombinant	Polymersomes derivative of ELP	Amphiphilic; arginine (R) & glutamine (Q)	$(\text{R}_5\text{Q}_5)_2\text{-F}_{20}$	SV: 20-200 nm GV: 1-20 μm	Controlled formation of peptide vesicles	98

H ₄₀ I ₃₀	Recombinant	ELP vesicle from "prebiotic amino acid"	Amphiphilic diblock. Histidine (H) & Isoleucine (I)	(VPGHG) ₂₀ (VPGIG) ₃₀	>10 μm	Protein membrane-based compartments (PMBCs) encapsulating anabolic reactions and catalysis.	102
S ₂₀ I ₃₀	Recombinant	-	Serine (S) & isoleucine (I)	(VPGSG) ₂₀ (VPGIG) ₃₀	5 μm	-	102
H ₅ L ₄	Recombinant	-	Histidine (H) & lysine (L)	(VPGHG) ₅ (VPGLG) ₄	2 μm	-	102

*Natural domains (in block-polymerization) are reported to be composed through the conjugation of the polydisperse blocks with mixed amino acids to yield an average composition and no specific composition could be defined (ref. 26).

#POP(V)-25% (where 'V' designates the guest residue amino acid and 25% designates the fraction of oligoalanine) with ELP(V₄A₁)