Doctoral Dissertation
Doctoral Program in Physics (35thcycle)

Statistical physics of molecular sorting in living cells

By

Elisa Floris
******

Supervisor(s):
Prof. Andrea Gamba, Supervisor
Prof. Luca Dall'Asta, Co-Supervisor

Doctoral Examination Committee:
Dr. Andrea Chiariello, Referee, Università di Napoli Federico II, Italy
Prof. Matteo Osella, Referee, Università di Torino, Italy

Politecnico di Torino
2023

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

<div align="right">

Elisa Floris

2023

</div>

Abstract

Molecular sorting is a sophisticated mechanism exploited by eukaryotic cells to maintain their inner order and guarantee proper physiological functioning. By means of this mechanism, specific biomolecules dwelling on the outer plasma membrane and on the internal membranes are selectively concentrated into spatially localized domains and engulfed into submicrometric lipid vesicles, that are actively delivered to their right destinations. A similar sorting process is also involved in the assembly and budding of enveloped viruses. This thesis is devoted to the study of the statistical properties of molecular sorting investigated by means of an abstract model, that assumes that this non-equilibrium process emerges from the coupling of two main physical principles: (a) molecule phase separation into domains and (b) domain-induced membrane bending, leading to vesicle nucleation.

Contents

Chapter 1

Introduction

Compartmentalization is crucial for cell survival. The robust physiological functioning of eukaryotic cells relies on a intricate system of distinct inner structures, or compartments, each of which is designated to accomplish a specific task [1–4]. This functional diversification originates from the unique biochemical composition of each subregion. Indeed, each compartment is endowed with selected groups of specialized proteins and other auxiliary molecules that cooperate for a common purpose, defining its identity and function. Orchestrating the myriad of incessantly diffusing and interacting molecules, that need to be at the right place at a precise time, is thus vital for the cell and errors can lead to the development of severe diseases. How molecules are initially partitioned into different compartments and are maintained spatially separated are questions still under investigation.

A major mechanism for the organization of intracellular matter is the process known in the biological literature as molecular sorting, which takes place on the plasma membrane and on the membranes enveloping the inner cell compartments [5, 6]. By this process, each of the thousands of proteins in a cell is correctly assigned and delivered to its appropriate membrane region via small lipid vesicles through well-defined pathways. Although huge efforts in investigating this process have been made through the decades, the underlying physical mechanisms need to be further explored. In this regard, it has been recently proposed that this highly complex process may emerge from the coupling of two simple physical mechanisms [7]: a) phase separation of specific proteins into localized, homogeneous, submicrometric domains, and b) membrane bending induced by protein crowding, leading to the formation

of protein-enriched vesicles. The combination of these two mechanisms ultimately results in a spontaneous demixing process. Based on these assumptions, a non-equilibrium phenomenological theory of molecular sorting has been formulated [7], along with a microscopic lattice-gas model, describing molecule insertion on the lipid membrane, molecule diffusion and aggregation into domains, and molecule extraction as part of a domain once it has reached a characteristic extraction size.

The main goal of this thesis is to gain further insights into this fundamental process, firstly by exploiting the molecular sorting model developed in [7] and then by extending it to include more aspects.

A central role in the process of domain growth is played by the critical domain size which, as predicted by the theory of phase separation, distinguishes domains into subcritical domains with a low probability of reaching the extraction size, and supercritical domains which irreversibly grow and are ultimately extracted from the system. Sorting domains observed in experiments are commonly classified into two groups: productive domains, that evolve into vesicles, and unproductive domains, characterized by smaller size and shorter lifetimes, which rapidly dissolve. Therefore, it is tempting to interpret the different final fates a domain can undergo in terms of the critical domain size. Numerical results of the lattice-gas model of molecular sorting show good qualitative agreement with experimental data on domain lifetimes and sizes, supporting the hypothesis that sorting is driven by a phase separation process.

Thanks to recent advances in imaging techniques, it is now possible to directly visualize the subsequent events leading to the simultaneous sorting and dispatching of multiple species in living cells [8–11]. Here, the phenomenological theory of molecular sorting, which for simplicity was developed by considering a single molecular species [7], is thus extended to the case of the simultaneous sorting of a plurality of species. A combination of analytical estimate and numerical simulations shows that, in principle, a large number of species can be sorted in parallel on a membrane region without significant crowding effects.

An increasing amount of experimental studies shows that the valence, defined as the average number of interacting neighbors of a molecule in a domain, plays a crucial role in the formation of protein domains [12–14]. The effect of varying valence in this context is investigated by means of the lattice-gas model of molecular sorting, simulating protein diffusion and aggregation on triangular, square and hexagonal

lattices. Numerical results reveal that molecules with higher valence are sorted with greater efficiency.

Lastly, to explore the effect of the membrane bending on the sorting process, a simple model is introduced here, which simulates the sorting of a binary mixture of molecules onto a fluctuating, topologically-varying membrane lattice. Thus, differently from the previously introduced model [7] where the membrane was described as a static lattice of fixed size, here the lattice is allowed to undergo shape deformations and topological changes corresponding to events of fusion and fission of protein-loaded vesicles. The properties of the lipid membrane and of the detached vesicles are investigated in this new framework.

The organization of the thesis is as follows. In Chapter 2, the phase-separation-driven process of molecular sorting is introduced, highlighting its vital role in organizing intracellular matter. Basic concepts from the classical theory of phase separation and the theory of membrane elasticity are also recalled here. Chapter 3 contains the phenomenological theory of molecular sorting developed in [7], along with a description of its lattice-gas implementation. Chapters from 4 to 7 contain the results on which I have contributed: in particular, in Chapter 4 the role of the critical domain size in molecular sorting is studied, Chapter 5 concerns the process of sorting multiple molecular species in parallel, Chapter 6 investigates the impact of molecular valence on the efficiency of sorting, and in Chapter 7 a new model of molecular sorting explicitly accounting for the coupling of lipid membrane and proteins dynamics is proposed.

Chapter 2

Molecular sorting as a phase separation process

The content of this Chapter is partially based on the published paper [4].

In this Chapter I review basic notions about cell compartmentalization, phase separation processes and the physics of membranes.

2.1 Intracellular organization and molecular sorting

To effectively interact with their environment, cells have to generate inner structures by a sequence of symmetry-breaking events, whereby regions devoted to different vital functions emerge either spontaneously or under the influence of specific driving cues [15]. For instance, stem cells generate differentiated daughter cells by asymmetric cell division, endowing them with specific fate determinants [16] (Fig. 2.1(a)). Cells migrating in a chemotactic gradient are initially round but under the influence of a chemotactic factor develop a chemically differentiated front and back allowing them to migrate towards a chemoattractant source [17] (Fig. 2.1(b)). Epithelial cells mature chemically differentiated apical, basal and lateral region to integrate into the fine planar architecture of an epithelial tissue [18] (Fig. 2.1(c)). In these examples, the plasma membrane is subdivided in specialized, chemically differentiated domains by a selforganized phase separation process [19–22] which in its turn guides the asymmetric positioning of inner structures, such as the
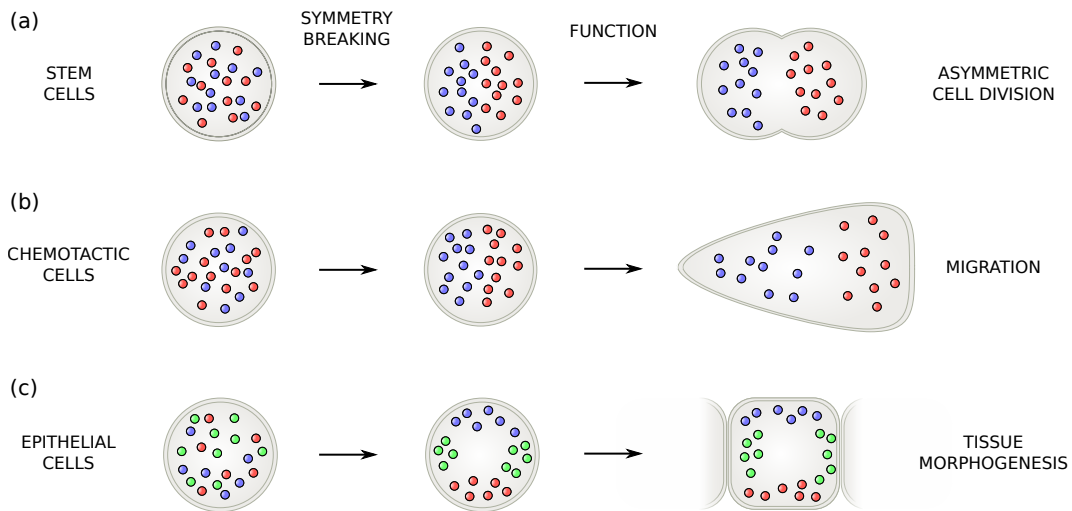
Fig. 2.1 Symmetry breaking and generation of cellular functions. (a) In asymmetric cell division, fate determinants are distributed asymmetrically between the mother and daughter cell [16]. (b) In chemotactic cells, different molecular factors accumulate in the growing anterior part and in the retracting posterior part of the cell [17]. (c) During epithelial tissue morphogenesis, the apical, lateral and basal regions of the cell acquire different molecular identities [18, 29].

cytoskeleton, and membrane-bound organelles, such as the Golgi and endoplasmic reticulum, in the intracellular space [23–26]. A sophisticated machinery of molecular sorting and dispatching contributes to the generation and maintenance of the symmetry-broken, polarized state [5, 27, 28]: the biochemical constituents of these localized domains dwelling on the plasma membrane and on the membranes of inner cellular bodies are engulfed in small lipid vesicles that, once formed, detach from the membrane and are subsequently delivered to their appropriate destinations.

Differently from purely chemical systems that may rely on 'passive', non-energy consuming intermolecular interactions for the generation and maintenance of separated phases, biological systems typically employ energy-consuming, non-equilibrium, 'active' processes to sustain heterogeneity among the distinct phases that identify its inner compartments [30]. In particular, the ability of biomolecules to separate into phases characterized by distinct states of matter, ranging from liquid to gas or solid, is fundamental to spatially segregate proteins and nucleic acids in cells, thus generating specific compartments with specialized functionalities. Diverse biological processes ranging from RNA metabolism, DNA damage response and

signal transduction exploit the fine levels of compartmentalization provided by phase-separated domains for their proper functioning [31–33].

Membrane-bound compartments

Cell membranes are primarily responsible for compartmentalization: each organelle is enclosed in its membrane, which serves as delimiting interface between the organelle's contents and the surrounding cytosol. In the intracellular space, membranes not only constitute boundaries but also, and more importantly, define each organelle's function. Indeed, apart for the common structure provided by a self-assembled double layer of lipids all cell membranes share, each membrane hosts a specific set of proteins and other macromolecules, determining its function.

Lastly, membranes play a pivotal role in diverse cellular processes, including molecular sorting and trafficking to/from the external cell environment and between different intracellular locations. Thus, membranes are required to continuously undergo dynamical changes and assume a broad diversity of morphological states, ranging from non-spherical shapes to spherical ones.

Membraneless organelles

Recent evidences have shown that phase separation is a key driver not only of the spatial organization of molecules on cell membranes, but also of the biogenesis of a large class of membraneless organelles contained in the cytoplasm and in the nucleoplasm. These organelles, although lacking an enclosing membrane which physically separate the internal content from the surrounding medium, maintain a coherent structure and are characterized by a specific biochemical composition [34, 35]. A far from exhaustive list includes the nucleolus, Cajal bodies, nuclear speckles, stress granules, P-bodies, and germ granules.

2.2   Phase separation in cell biology

Processes of phase separation have emerged in the recent decade as an ubiquitous feature of eukaryotic cell physiology. Phase separation is a natural process that the

cell may exploit to spatially localize biochemical reactions and cellular functions in appropriate subcellular structures and compartments. Two large classes of macromolecules have been identified, that exhibit a particular tendency to phase separate: proteins and RNA molecules containing several repeats of similar binding sites and/or weak interaction regions (multivalent molecules) and proteins (in particular, RNA- and DNA-binding proteins) containing stretches of unstructured, partially unfolded regions enriched in particular polar and charged amino acids (intrinsically disordered, or prion-like, protein regions) [12, 32, 36]. Favored by negligible activation barriers, such feeble interactions promote spontaneous and reversible organization of macromolecules in phase-separated domains, often referred to also as 'droplets', or 'condensates' [37]. In most cases the state of aggregation of such domains has been reported to be liquid-like, or gel, but occasionally also solid domains have been observed, mainly in association with pathological conditions [32]. On lipid membranes, phase separation into localized domains enriched in specific molecular factors can be also driven by the indirect intermolecular interactions established by networks of autocatalytic loops. Such 'soft' domains may have a gas-like structure, as their constituents do not need to be constantly in direct contact [19, 38]. The progressive 'coarsening', or 'ripening' of phase-separated domains evidenced by time-lapse experiments, which is a signature of phase separation in the presence of a finite pool of molecular factors, has been observed in the cytosol, nucleus, and on lipid membranes, both in vitro (Fig. 2.2(a, b)) and in vivo (Fig. 2.2 (c-f)). Computer simulations of quantitative models of phase separation reproduce the observed coarsening dynamics (Fig. 2.2 (g, h)).

By organizing biochemical reactions in time and space, phase-separated domains allow them to proceed at the right pace by preventing undesired side-reactions in multistage processes [31, 33], storing away biomaterials when the cell comes under stress and releasing them gradually in normal conditions [36, 43], selectively inhibiting or promoting reactions by sequestering [44] or concentrating reactants [45, 46].

Importantly, at the larger, cellular scale, the regulated formation of phase-separated domains allows the right processes to take place at the right time and in the right place. A far from exhaustive list of examples is briefly reviewed here. The centrosome, which serves as the main microtubule organizing center during mitosis, is nucleated by the centriole and grows by a phase-separation process [47]. Then, by absorbing tubulin from the cytosol and concentrating it in its interior, it favors the
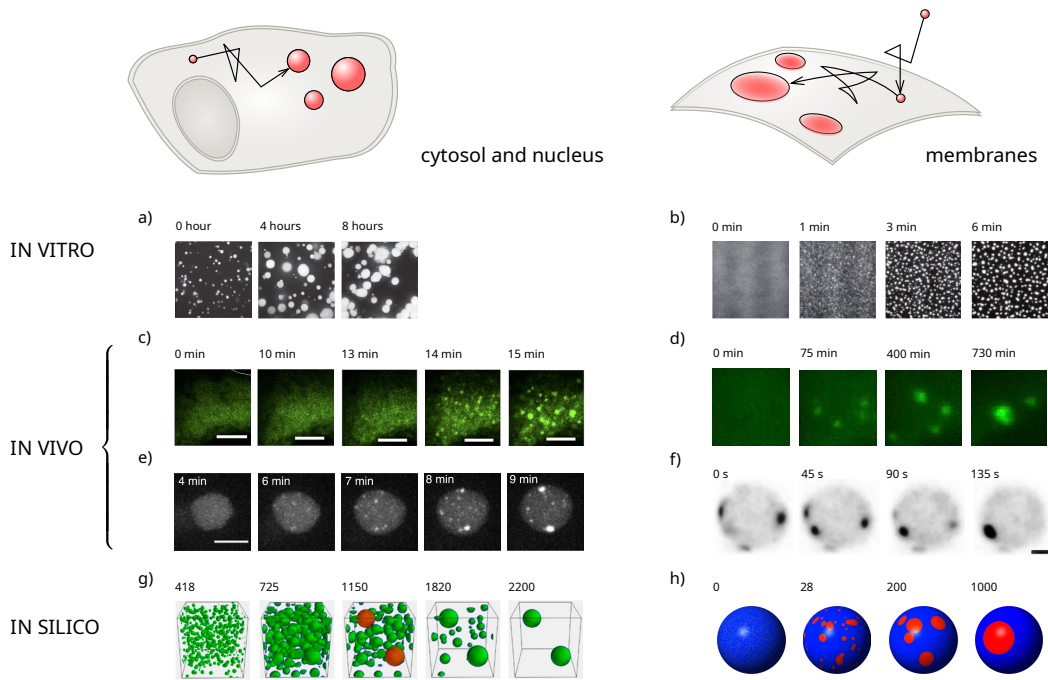
Fig. 2.2 Nucleation and coarsening of selforganized domains in biological phase separation (reproduced from the original with permission, in modified form): (a) prion-like FUS protein, associated with the neurodegenerative disease ALS [32]; (b) LAT protein, taking part in T cell receptor signal transduction [33]; (c) stress granules [39]; (d) post-synaptic densities [40]; (e) nucleoli and extranucleolar droplets [41]; (f) polarity establishment in yeast [42]; (g) simulation of the coarsening kinetics of nucleoli and extranucleolar droplets[41]; (h) simulation of coarsening kinetics in the establishment of cell polarity [20, 22].

multiple nucleation of microtubules from its surface [48]. The nucleolus, which is responsible for ribosome biogenesis, is nucleated around a specific ribosomal DNA region and grows by phase separation, as evidenced by the observed coarsening kinetics (Fig. 2.2e). Germ granules, aggregates of protein and RNA that determine the differentiation fate of daughter cells generated by asymmetric cell division, form during mitosis by a phase separation process [49], which is asymmetrically driven by the previous formation of opposite phase-separated domains on the plasma membrane of the mother cell [50].

2.3   Molecular interactions inducing phase separation

The interaction between homotypic molecules leading to their phase separation into distinct sorting domains may be either *direct*, such as in the case of the weakly adhesive electrostatic interactions between unstructured molecule regions involved in biological liquid-liquid phase separation [37], or *indirect*, as in the case of the effective, contactless interactions induced by enzyme-driven feedback loops involving lipid and molecules. The latter, indirect interactions have the potential to induce diffusion-limited phase separation, originally studied in the context of cell polarity [19, 51, 22, 3, 52], and are involved in molecular sorting processes where the segregation of distinct molecular species in separate sorting domains is not controlled by direct homotypic intermolecular interactions [53].

Direct interactions

Mixing-demixing transitions have been thoroughly studied in Physics [54–56]. There, the ordering of similar molecules by mutual affinity in spatially separated domains is driven by attractive or repulsive interactions, primarily of electrostatic origin, such as those emerging from the interaction of permanent or induced dipoles [57–59]. The formation of phase-separated domains becomes possible when the demixing tendency of mutually attractive interactions outcompetes the tendency of thermal agitation to mix and homogenize the molecular components of a system. For this reason, in the simplest systems the mixing-demixing transition may take place abruptly when some control parameter (such as the temperature, or the concentration of a particular component) crosses some critical value, and a tipping point is reached.

The initiation of this process of separation into different phases, characterized by distinct physical properties and/or molecular compositions, can occur spontaneously as a consequence of stochastic fluctuations. For instance, the spontaneous nucleation of a germ of a (solid, liquid or gel) condensate phase from a solution starts from the random encounter of two molecules of the solute, resulting in the formation of a (stable or transient) dimer. The growth of a small condensate domain of molecules of the solute is then driven by the balance between the influx of molecules from the solution and the 'evaporation' of molecules from the domain [60]. If, after the initial nucleation stage, phase separation takes place sufficiently close to thermodynamic equilibrium, it is possible to describe it in thermodynamical terms as a competition

between a bulk and interface free energy. Here, the growth of a spherical domain of the condensate phase allows the system to lower its energy proportionally to the increase in the volume of the domain itself, while at the same time, the free energy of the system is raised by an amount proportional to the increase of interface area, where the energetically favorable contacts are not saturated (the proportionality coefficient is called a 'surface tension'). A tug of war therefore arises between these two competing effects.

Several observable effects derive from this scenario [55, 60–63]:

1. Phase separation takes place via a *switch-like onset* when the concentration of the solute exceeds the threshold concentration that allows a large condensate domain to coexist at equilibrium with the solution.

2. The process tends to *minimize the interface area* (or *perimeter length* in two-dimensional systems) between the two phases, leading to the formation of approximately spherical (or circular, in two-dimensional systems) growing domains.

3. The speed of the process is controlled by a *degree of metastability*, which is approximately proportional to the difference between the concentration of the solute and the threshold concentration.

4. There is a *critical size* under which domains are unstable and tend to disappear, since for small domains, the energy advantage coming with an increased volume does not repay the cost of an extended droplet boundary; the critical size of approximately spherical or circular domains is inversely proportional to the degree of metastability.

5. Stable domains (i.e., domains larger than the critical size) can be generated either by a large enough random fluctuation (*homogeneous nucleation*), or by the creation of a large enough nucleation center by some external action (*heterogeneous nucleation*).

6. When competing for a limited pool of molecules, domains undergo competitive growth, also known as *coarsening*: larger domains grow at the expense of the molecules that 'evaporate' from smaller ones, so that at equilibrium a unique domain survives, in a sort of winner-take-all mechanism; moreover, the growth of domains becomes slower and slower with time, as long as the

solute is sequestered by the growing domains, and as a consequence, the degree of metastability decreases and the critical size grows; Lifshits and Slyozov [64, 62] found that under these conditions the average size of domains grows as $(\text{time})^{1/3}$ in both two- and three-dimensional environments. This is ultimately a consequence of the fact that molecules diffuse in a restricted environment without being created or destroyed (in this case it is said that the system exhibits a *locally conserved* order parameter [55]).

A similar scenario is realized in the case of an initially well-mixed binary mixture of a population of molecules of types A and B (Fig. 2.3(a)), characterized by homotypic intermolecular affinity, where islands of pure A- and B-phases can be nucleated and grow in the 'sea' of the well-mixed phase.

In both cases, a mathematical description of phase separation can be given by introducing a coarse-grained *order parameter* ϕ (that in the case of the binary mixture can for instance be thought as the difference between the local concentrations of the A- and B-molecules) and a Landau-Ginzburg free-energy density:

$$f = \frac{K}{2}|\nabla\phi|^2 + V(\phi) \tag{2.1}$$

where the gradient term penalizes the interfaces between the two phases (K is thus proportional to the surface tension), while $V(\phi)$ is a potential, which depends on the strengths of the intermolecular interactions. There exist a region of parameter space (Fig. 2.3(b)) in which the coexistence of competing phases is possible, as $V(\phi)$ develops two minima (Fig. 2.3(c)), corresponding respectively to each of the two physically realizable phases. If $V(\phi)$ is symmetric, both phases are globally stable, otherwise one of the two phases is metastable.

Indirect interactions

Classical theories of phase separation rely on concepts of equilibrium statistical mechanics, such as the free energy, that in some cases can be used to approximately describe also non-equilibrium dissipative processes during their relaxation toward an equilibrium state. However, many of the inner workings of living cells are intrinsically nonequilibrium, since they are continuously driven by external and/or
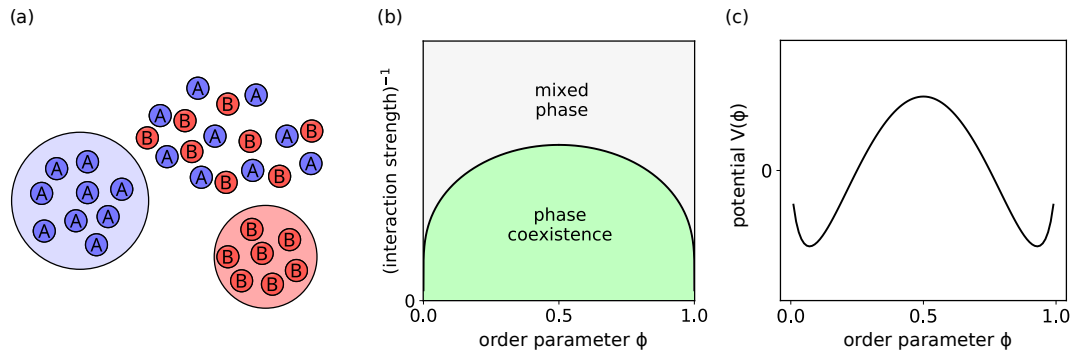
Fig. 2.3 Mixing-demixing transition. (a) Schematic representation of the formation of phase-separated domains in a 'sea' of the mixed phase. (b) Phase diagram of the process. (c) Potential part of the free energy in the phase-coexistence region, with the characteristic bistable shape; the two potential wells correspond to the two stable phases.

internal forces, and can involve transport phenomena and enzymatic processes in which individual molecular components constantly and irreversibly consume and dissipate energy, a defining feature of 'active' matter [30, 65, 66]. Such intrinsically out-of-equilibrium processes permit the existence of a larger variety of stationary states than those realizable in close-to-equilibrium conditions [47, 67]. Such processes contribute to generate and maintain the symmetry-broken, structured, compartmentalized state which allows the cell to perform its complex, vital functions. An example of such order-generating, energy-consuming nonequilibrium process is provided by the phase separation of cell membranes into polarized signaling domains driven by autocatalytic loops [19, 68–71, 53]. In this context, the attractive or repulsive interaction between homotypic molecules is often not direct, but effectively mediated by autocatalytic feedback loops involving auxiliary molecules and sustained by a continuous energy influx, that may be provided for instance by ATP hydrolysis and is consumed by individual catalytic events.

A simple model of this active phase separation process describes a population of molecules of types A and B, bound to the lipid membrane, where they can laterally diffuse with diffusivity constant D. The two molecular types are interconverted by the action of two enzymes E_A, E_B that shuttle between the lipid membrane and the cytosol, where they rapidly diffuse. Each of the two enzyme types binds preferentially to membrane regions enriched in their own product, thus realizing a simple reinforcing feedback loop (Fig. 2.4) [19, 20, 51, 22, 72]. Based on some of
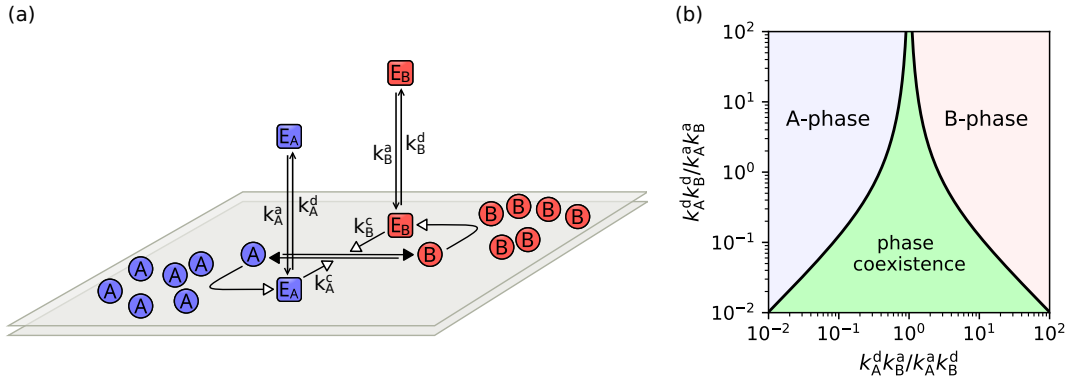
Fig. 2.4 (a) Abstract model for active phase separation on lipid membranes [19, 20, 51, 22]. (b) Phase diagram. Concentrations are measured in units of $A^{\text{tot}} + B^{\text{tot}}$; the graph is symmetric with respect to the vertical axis centered in $k_A^c E_A^{\text{tot}} / k_B^c E_B^{\text{tot}}$ (here assumed to be unity).

its characteristic properties, this kind of model has been termed 'mass-conserved', 'bistable', or 'wave-pinning' by different authors [73–75].

A mathematical description of this process can be given by considering as an order parameter the difference in the surface concentration of the molecules of types A and B [19, Supp. Text]. The evolution of ϕ in time can be reduced to the minimization of the effective free energy density:

$$f_{\text{act}} = \frac{D}{2} |\nabla \phi|^2 + V_{\text{act}}(\phi), \tag{2.2}$$

where $V_{\text{act}}(\phi)$ is an effective potential, which depends on the chemical reaction rates indicated in Fig. 2.4(a). In the region of parameter space depicted in light green in Fig. 2.4(b) the coexistence of competing phases is possible, as $V_{\text{act}}(\phi)$ develops two minima, corresponding respectively to each of the two (A- and B-enriched) phases. This is analogous to the mathematical description of the classical phase separation of a binary mixture, except that $V_{\text{act}}(\phi)$ now depends on the kinetic rates of a set of non-equilibrium autocatalytic reactions, instead of the equilibrium strength of direct intermolecular interactions. A similar mathematical structure implies similar properties: domains of pure A- and B-phases may nucleate and grow, exhibiting all the 'classical' effects, including switch-like onset, minimization of surface tension, existence of critical size, and coarsening [19, 20, 51]. Interestingly, in these conditions coarsening is faster than in the classical Lifshitz-Slyozov prediction, since the order parameter is not conserved: molecules may 'evaporate' from any part

of a domain, diffuse rapidly in the cytosol, and be captured again in another point of the membrane; correspondingly, the average size of domains is predicted to grow as $(\text{time})^{1/2}$ [20, 51].

These selforganized 'active' domains exhibit peculiar features. They cannot exist at the stationary state without a constant influx of energy, and are therefore intrinsically out of equilibrium. Since direct homotypic molecular interactions are here substituted by effective interactions mediated by the autocatalytic loops, the intermolecular distance in such domains can be larger than the molecular size, a characteristic that can be assimilated to that of a 'gas' phase. In principle, such domains may be expected to cover large extents of space and to exhibit a high degree of plasticity and fast recycling of their constituents. Interestingly, the size of these membrane domains is limited by the depletion of their constituents from the cytosolic reservoir, and can therefore in principle be controlled by regulating their cytosolic concentrations [19, 20, 51].

It is also worth observing here that actual selforganized active domains are driven by selfreinforcing catalytic loops that usually involve a chain of events where a multiplicity of molecular species participate. Therefore, such domains are expected to be typically multicomponent, and to host a spatially-localized, higher-than-average concentration of the 'clan' of all of the molecular species that take part into the relevant autocatalytic feedback loops (the red and blue 'clans' in Fig. 2.4 can be seen as an abstract example).

2.4   Kinetics of phase separation

Phase separation corresponds to the decay of a thermodynamic system from a metastable state (corresponding to a local minimum of the free energy) to a stable one (global minimum). This transition occurs through the subsequent steps of nucleation, coarsening, and coalescence. The *nucleation* stage consists in the formation of tiny islands (called nuclei or droplets) of the stable phase in a "sea" of the unstable phase as a consequence of a large enough thermal fluctuation or driven by preexisting seeds of the new phase. The second stage is characterized by the growth of larger domains, or *coarsening*, at the expense of smaller ones, resulting in the decrease of the average number of domains. In the final stage phase coexistence is reached: two

large complementary domains remain in the system. In the following we will use the example of liquid-gas coexistence.

Nucleation

A simple description of the process of droplet nucleation is provided by the critical droplet theory [54, 76]. It relies on a quasi-equilibrium assumption, according to which a metastable state can be treated as an equilibrium state if the lifetime of the metastable state is much larger than the relaxation time the system needs to reach that state. Using concepts from the thermodynamic fluctuation theory, the equilibrium probability describing the formation of a cluster composed of n particles (n-cluster) can be written as

$$f_{eq}(n) \propto e^{-\beta W_{min}}, \tag{2.3}$$

where W_{min} is the minimum reversible work necessary to form a n-cluster in the gas phase at pressure p_g and temperature T. If the process takes place at constant T and p_g, it can be shown that W_{min} is equal to the variation of the Gibbs free energy ΔG. This can be calculated as follows:

$$\begin{aligned} \Delta G &= \Delta F(T,V) + p_g \Delta V_g + p_g V_l \\ &= -n\Delta\mu + \gamma A(n) \\ &= -n\Delta\mu + \gamma S_d r^{d-1} n^{\frac{d-1}{d}}, \end{aligned} \tag{2.4}$$

where ΔF is the variation of the Helmholtz free energy, ΔV_g is the gas volume change, V_l is the volume of the liquid droplet, and $\Delta\mu = \mu_g(p_g) - \mu_l(p_g)$ is the chemical potential difference. The last line is obtained by using the capillarity approximation which consists of the two following assumptions:

i) the droplet is a sphere of incompressible liquid of radius $R = rn^{1/d}$ in d dimensions, with r being the radius of a single particle and d the dimensionality;

ii) the droplet surface energy is given by $U = \gamma A$, where $\gamma = \gamma_\infty(T)$ is the planar surface tension and $A = S_d r^{d-1} n^{\frac{d-1}{d}}$ denotes the surface area, with $S_d = 2\pi^{(d+1)/2}/\Gamma((d+1)/2)$.

For small n, the formation of a small cluster is energetically unfavourable due to the positive surface term, while for large n the bulk contribution dominates. The

maximum value of $\Delta G(n)$ is obtained in correspondence of the critical number of particles n_c, or analogously for the critical radius

$$R_c = \frac{d-1}{d} \frac{S_d r^d \gamma}{\Delta\mu}. \tag{2.5}$$

$\Delta G(R_c)$ is the nucleation barrier, that is the energy barrier a system has to overcome to form a new stable phase. On average, for $R > R_c$ the droplet grows, while for $R < R_c$ the droplet shrinks. $\Delta\mu$ appearing in the expression of the critical radius (Eq. 2.5) can be calculated by using the condition that the critical droplet is in equilibrium with the surrounding gas and the condition of liquid incompressibility. By expanding the free energy as a function of the radius around the critical region

$$\Delta G(R) \approx \Delta G(R_c) + \frac{1}{2} \frac{d^2 \Delta G(R)}{dR^2}\Big|_{R=R_c} (R-R_c)^2, \tag{2.6}$$

it follows that the equilibrium probability of finding a cluster of radius R can be expressed as

$$f_{eq}(R) \propto e^{\frac{4\pi\gamma}{k_B T}(R-R_c)^2}. \tag{2.7}$$

Recalling that, for large n, ΔG is dominated by the negative bulk contribution, it follows that $f_{eq}(n)$ diverges for $n \to \infty$, while it should remain finite for any n and t. To describe the dynamical nature of the nucleation-growth process, two kinetics methods were developed independently by Becker and Döring [77] and by Zeldovich [78], relying on a time-dependent cluster size distribution $f(n,t)$.

Coarsening

In the case of low supersaturation, the growth of a droplet is mainly due to the attachment of single diffusing molecules (diffusion-limited growth). Consider a droplet of radius R surrounded by a gas of non-interacting molecules at pressure p_g in three dimensions. The rate at which diffusing molecules hit the absorbing droplet is determined by the reaction rate theory [63]. Laplace equation with appropriate boundary conditions describes the stationary density profile of the gas n surrounding the droplet

$$D\nabla^2 n = 0, \tag{2.8}$$

with diffusion constant D and $n(\infty) = n_g$ and $n(R) = n_{0,R}$. Since the general solution of Laplace equation for $d \geqslant 3$ is $n(r) = A + Br^{2-d}$, one obtains

$$n(r) = n_g - (n_g - n_{0,R}) \left(\frac{R}{r}\right)^{d-2}. \tag{2.9}$$

The reaction rate K is given by the flux of particles across the surface of the droplet

$$\begin{aligned} K &= D \int_\Sigma \nabla n \cdot \hat{n} d\Sigma = S_d R^{d-1} D \frac{dn}{dr}\bigg|_{r=R} \\ &= S_d D(d-2) R^{d-2} (n_g - n_{0,R}), \end{aligned} \tag{2.10}$$

where $S_d = 2\pi^{(d+1)/2}/\Gamma((d+1)/2)$. Using the Gibbs-Thomson relation

$$n_{0,R} \simeq n_{0,\infty} \left[1 + \frac{2\gamma_\infty}{n_l k_B T} \frac{1}{R}\right], \tag{2.11}$$

for $d = 3$ one obtains

$$K = 4\pi DR \left[(n_g - n_{0,\infty}) - \frac{2\gamma n_{0,\infty}}{n_l k_B T} \frac{1}{R}\right]. \tag{2.12}$$

Calculating the critical radius R_c for $d = 3$ from Eq. 2.5 and using the following relation for the chemical potential difference

$$\begin{aligned} \Delta\mu &= \mu_g(p_g) - \mu_l(p_g) = k_B T \log \frac{p_g}{p_{sat}} \\ &= k_B T \log \frac{n_g}{n_{0,\infty}} \approx k_B T \frac{n_g - n_{0,\infty}}{n_{0,\infty}}, \end{aligned} \tag{2.13}$$

with p_{sat} is the saturation pressure, the absorption flux, defined as the number of particles per unit surface, is given by

$$i_0 = \frac{D}{R^2}(R - R_c)(n_g - n_{0,\infty}). \tag{2.14}$$

2.5   Membrane fluctuations

Molecular sorting is mediated by vesicles, small spherical membrane structures that bud off from a membrane region and fuse with another one, releasing their

content. Introducing a local curvature is a key requirement for initiating the formation of a vesicle. What are then the special strategies cells employ to easily deform membranes? In the last decades several studies have revealed that interactions between proteins, the second major component of membranes, and the lipid bilayer are the main drivers of membrane bending [79]. Lipid-protein interplay arises from the joint effort of distinct physical mechanisms [80, 81]: generation of an area mismatch due to the insertion of a helix into one leaflet of the membrane, scaffolding of the membrane, protein repulsion due to steric interactions [82–84], and protein phase separation [85]. A comprehensive understanding of all these mechanisms is still lacking.

Curvature

A membrane can be viewed as a two-dimensional surface embedded in a three-dimensional space (thin-sheet approximation). Consider, for instance, a vesicle of size 100 μm and thickness $3 - 5$ nm: the lateral extent of the membrane is much greater than its thickness [86]. Thus, notions of differential geometry can be used to describe a membrane and, in particular, its shape can be quantified in terms of the curvature. This is a local geometrical property which, for a planar curve, is defined as the inverse of the radius of the circle that best approximate the curve at each point (osculating circle). In the case of a surface, two circular mutually perpendicular arcs with radii R_1 and R_2 need to be defined. Their reciprocal $C_1 = 1/R_1$ and $C_2 = 1/R_2$ are called principal curvatures. The mean curvature is defined as

$$H_C = \frac{1}{2}(C_1 + C_2) \tag{2.15}$$

and the Gaussian curvature is defined as

$$H_G = C_1 C_2. \tag{2.16}$$

The Helfrich Hamiltonian

The Helfrich theory, widely employed to study the behaviour of biological membranes, is based on a continuum-mechanical approach. The use of a continuum description is justified by the two following properties: a) the area per lipid molecule in the

membrane is nearly constant; b) the fluctuation in the thickness of a lipid bilayer is negligible compared to its average thickness. In contrast to atomistic and coarse-grained models useful to study the biochemistry of interactions, but computationally expensive, a continuum framework allows to investigate the biophysical properties and their implications relevant to a cellular scale.

Owing to the thin-sheet approximation, the theory of membrane elasticity, independently developed by Canham [87] and Helfrich [88], can be considered as a special case of the theory of bent plates. The Helfrich Hamiltonian [88] has the form

$$H_{el} = \int dS \left[\frac{\kappa}{2} (H_C - H_0)^2 + \bar{\kappa} \, H_G \right], \tag{2.17}$$

where κ is the bending rigidity of the membrane, $\bar{\kappa}$ is the Gaussian rigidity, H_C is the local mean curvature, H_0 is the spontaneous curvature, assumed to be isotropic here, and H_G is the Gaussian curvature. The moduli κ and $\bar{\kappa}$ have the dimension of energy. Since a biological membrane is a self-assembled system, the bending modulus κ is of the order of $k_B T$. For closed surfaces, the Gaussian contribution is usually neglected since it is constant according to the Gauss-Bonnet theorem. In Eq. 2.17 a term corresponding to the surface tension is also neglected [89].

Chapter 3

Theory of phase-separation driven molecular sorting

The content of this Chapter is based on the published paper [90] and on the paper under review [91].

Molecular sorting is a major process responsible for the organization of cellular matter in eukaryotic cells [5]. This highly complex task is accomplished by selectively concentrating and distilling specific proteins and lipids that dwell on the plasma membrane and on the membranes of inner cellular bodies into submicrometric lipid vesicles. Once formed, these vesicles detach from the membrane and are subsequently delivered to their appropriate destinations. A similar process of sorting of specific molecular factors into localized membrane domains is involved in the assembly of enveloped viruses (such as HIV, SARS-CoV, and influenza) and in their budding from host cells [92–95].

Experimental evidences suggest that the self-organization of specific molecules into localized domains on cell membranes is the result of a phase separation process [72, 3, 96, 35]. Moreover, the formation of these domains has been shown to precede and initiate vesicle nucleation [97, 83, 98, 84, 85]. Starting from these observations, recently, a minimal model of molecular sorting has been developed, based on the idea that this sophisticated process may emerge from the combination of two fundamental physical mechanisms [7]: a) phase separation of specific molecules into localized sorting domains, and b) domain-induced membrane bending, leading

to the formation of vesicles constitutively enriched in the biochemical factors of the engulfed domains, thus resulting in a natural distillation process.

In Sec. 3.1 the phenomenological theory of molecular sorting introduced in [7] and extended in [91] is described, while in Sec. 3.2 the hard-core lattice-gas implementation of the model is explained.

3.1   Phenomenological theory

The phenomenological theory of phase-separation driven molecular sorting introduced in Ref. [7] is here extended to the case of $N > 1$ non-interacting species, based on the following non-equilibrium steady-state picture. A constant flux $\phi = \sum_{i=1}^{N} \phi_i$ of "sortable" cargo molecules is deposited on the lipid membrane at random positions, where ϕ_i denotes the average number of particles of the i-th species. Molecules diffuse laterally and attractive (direct or indirect) interactions between homotypic molecules (i.e., molecules belonging to the same species) can lead to the formation of multiple sorting domains enriched in the molecules of a particular species, thus inducing a natural *demixing* process. Since domain formation is characterized by competing effects, according to classical nucleation theory, a critical size A_c is required for a domain to continue to grow irreversibly and avoid decay (Sec. 2.4 and Refs. [77, 78, 60]). Once formed, sorting domains coarsen due to the incoming flux of laterally diffusing molecules, and are eventually extracted from the membrane in the form of lipid vesicles of characteristic area $A_E = mA_0$, with $A_E \gg A_c$, naturally enriched in the molecules of the engulfed sorting domain. It follows that the growing domains coexist with a continuously repleted two-dimensional "gas" of laterally diffusing molecules in a statistically stationary state, whose properties are determined by the incoming flux of molecules and by the *effective interaction strength* which measures the interactions between homotypic molecules, irrespective of the microscopic (direct or indirect) origin of the attraction. A schematic representation of the minimal model of molecular sorting is shown in Fig. 3.1.

It is important to consider here that in biological phase separation, the formation of localized domains is often driven by the action of positive feedback loops that involve several molecular factors (Sec. 2.2). The resulting phase-separated domains are thus enriched in the whole set of molecular factors that participate in each feedback loop. Distinct feedback loops lead to the formation of localized domains
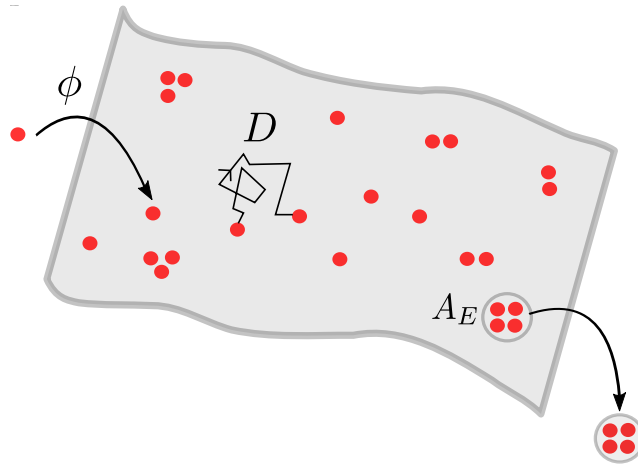
Fig. 3.1 Schematic representation of the minimal model of molecular sorting: cargo molecules are inserted at random positions on the lipid membrane with a constant flux ϕ, diffuse laterally with diffusivity D, and can aggregate into domains, which are extracted from the membrane if their size reaches a characteristic value A_E.

endowed with distinct chemical identities. To simplify the discussion, we consider here a *coarse-grained* description, where each of the distinct chemical identities emerging from phase separation is represented by a single, *representative* molecular species, such as a well-defined cargo molecule. It should be kept in mind, however, that on a real cell membrane, a domain enriched in the representative molecular species will also be enriched in a whole clan of associated molecular factors. For instance, the endocytic sorting of a cargo molecule also involves the accumulation in sorting domains of a whole set of auxiliary molecules, playing the roles of scaffolds, regulators, etc. [5, 99]. In the same spirit, we consider here *effective* attractive interactions between members of homotypic representative species, that may result from the participation of the representative species, together with a clan of associated molecular factors, to a common network of reinforcing feedback loops, even in the absence of any *direct* homotypic interaction.

Domain size distribution

In the low density regime, the process of formation of domains enriched in a specific type of cargo molecule is approximately independent of the formation of domains of the other species. If we consider a region of linear size L of the order of the

average interdomain half distance, centered around a growing supercritical domain of approximately circular shape and radius R, the quasi-static profile $n_R(r)$ of the density of the gas of freely diffusing molecules in the proximity of the domain can be approximately obtained by solving a Laplace equation with Dirichlet boundary conditions $n_R(R) = n_0$ and $n_R(L) = \bar{n}$ (following the same reasoning of Sec. 2.4), obtaining

$$n_R(r) = n_0 + \frac{\log(r/R)}{\log(L/R)}\Delta n, \tag{3.1}$$

where $r \geq R$ denotes the distance from the domain center, and $\Delta n = \bar{n} - n_0$. Domain growth is induced by the flux Φ_A of molecules from the gas to the domain, which can be calculated by integrating the flux density $-D\nabla n_R(r)$ across the boundary of the domain of size $A = \pi R^2$, obtaining

$$\Phi_A = \frac{4\pi D \Delta n}{\log(A_L/A)}, \tag{3.2}$$

where D is the lateral diffusivity of the molecules and A_L is a circular region of size $A_L = \pi L^2$. This formula implies that the domain will grow according to the dynamic equation

$$\dot{A} = \Phi_A A_0 = \frac{4\pi A_0 D \Delta n}{\log(A_L/A)}. \tag{3.3}$$

In a membrane system where sorting domains may be assumed to be approximately evenly distributed, the statistics of supercritical domains can be conveniently described in terms of the number density $\mathcal{N}(t,A)\,dA$, giving the average number per unit membrane area of supercritical domains with size comprised between A and $A + dA$. Since the effects of random fluctuations can be approximately neglected in the case of supercritical domains, $\mathcal{N}(A,t)$ satisfies the continuity equation

$$\frac{\partial \mathcal{N}}{\partial t} + \frac{\partial}{\partial A}(\dot{A}\mathcal{N}) + \gamma(A)\mathcal{N} = 0, \tag{3.4}$$

where the rate of removal of domains of size A from the system is $\gamma(A) = 0$ for $A < A_E$, and $\gamma(A) = \gamma_0 > 0$ for $A > A_E$. The stationary solution of Eq. (3.4),

$$\mathcal{N}_{\mathrm{st}}(A) = \frac{J\log(A_L/A)}{4\pi D \Delta n}\exp\left[-\int_{A_c}^{A}\frac{\gamma(a)\log(A_L/a)}{4\pi A_0 D \Delta n}\,da\right] \tag{3.5}$$

has a universal logarithmic behavior for $A < A_E$. The normalization constant J can be determined from the steady state condition

$$\phi = \int_{A_c}^{\infty} \Phi_A \mathcal{N}_{st}(A)\, dA \simeq J A_E \tag{3.6}$$

for large γ_0 and $A_E \gg A_c$. Assuming that the incoming flux ϕ of molecules is evenly distributed in average among all available supercritical sorting domains, and neglecting logarithmic corrections, the average total number of supercritical domains per unit area is given by

$$\bar{N}_d \sim \frac{\phi}{\Phi_A} \sim \frac{\phi}{D \Delta n}. \tag{3.7}$$

Optimality

In the non-equilibrium stationary state, the average number ϕ_i of particles of the i-th species injected into the membrane system per unit time and unit area equals the analogous number of such particles leaving the system as a consequence of domain extractions. Under the assumption that supercritical domains grow irreversibly until extraction, one gets

$$\phi_i = m_i \frac{d\bar{N}_{d,i}}{dt}, \tag{3.8}$$

where $d\bar{N}_{d,i}/dt$ is the rate of formation of supercritical domains of the i-th species per unit membrane area, and $m_i = A_E/A_{0,i}$, with $A_{0,i}$ the average area occupied by a sorted molecule in a sorting domain. The rate of formation of such domains depends on the frequency of formation of germs of new sorting domains and on the probability that those germs reach the supercritical stage. Since the formation of a domain is initiated by the encounter of two freely diffusing molecules of the i-th species, the rate of domain formation has to be proportional to the square of the average free molecule density $\bar{n}_i$ [7]. Therefore, the rate of formation of supercritical domains can be expressed phenomenologically as

$$\frac{d\bar{N}_{d,i}}{dt} = C_i D_i \bar{n}_i^2, \tag{3.9}$$

where C_i is a dimensionless quantity representing the macroscopic, effective strength of the attractive interaction acting between homotypic molecules [7, 90]. According to a general steady-state relation valid for open systems in a driven non-equilibrium

stationary state, the average density of particles in the system is given by the product of the average density flux of particles and the average residence time of a particle in the system [100]. In the present case, this relation can be applied to several entities that populate the membrane in the statistically stationary state. For the total, average density ρ of molecules (both freely diffusing and bound to sorting domains) of all N species in the stationary state one has

$$\rho = \phi \, \bar{T}, \qquad (3.10)$$

where $\bar{T}$ is the average molecule residence time on the membrane, and $\phi = \sum_{i=1}^{N} \phi_i$. This shows in particular that for fixed values of the molecular flux ϕ, the average residence time $\bar{T}$ is simply proportional to the average molecule density ρ. For the average density of freely diffusing molecules of the i-th species one finds

$$\bar{n}_i = \phi_i \, \bar{T}_{f,i}, \qquad (3.11)$$

where $\bar{T}_{f,i}$ is the average time a molecule of the i-th species spends in the gas. For the average density of supercritical domains of the i-th species one has

$$\bar{N}_{d,i} = \frac{d\bar{N}_{d,i}}{dt} \, \bar{T}_{d,i} = \frac{\phi_i}{m} \, \bar{T}_{d,i}, \qquad (3.12)$$

where $\bar{T}_{d,i}$ is the average lifetime of a sorting domain. The latter, in the limit of approximately absorbing domains, is of the order of the average time a molecule of the i-th species spends as a part of a sorting domain. For simplicity, we only analyze here the symmetric case, where $A_{0,i} = A_0$, $C_i = C$ and $D_i = D$ for all $i = 1, \ldots, N$, and also assume that $\bar{T}_{f,i} = \bar{T}_f$, $\bar{T}_{d,i} = \bar{T}_d$, and $\phi_i = \phi/N$ for all $i = 1, \ldots, N$. The total density of molecules in the gas is then

$$\bar{n} = \phi \, \bar{T}_f, \qquad (3.13)$$

while the total number of supercritical domains per unit area is

$$\bar{N}_d = \phi \, \bar{T}_d / m. \qquad (3.14)$$

Combining the relations (3.8–3.14), all the main quantities describing the behavior of the system in the non-equilibrium, statistically stationary state can be explicitly expressed in terms of the number of species N, the total incoming molecule flux

ϕ, the extraction size $m = A_E/A_0$, the diffusivity D, and the phenomenological interaction strength C, as:

$$\bar{n} \sim \left(\frac{\phi N}{mCD} \right)^{1/2}, \tag{3.15}$$

$$\bar{N}_d \sim \left(\frac{mC\phi N}{D} \right)^{1/2}, \tag{3.16}$$

$$\bar{T}_d \sim \left(\frac{m^3 CN}{D\phi} \right)^{1/2}, \tag{3.17}$$

$$\bar{T}_f \sim \left(\frac{N}{mCD\phi} \right)^{1/2}. \tag{3.18}$$

The efficiency of the molecular sorting process in the steady state is inversely proportional to the mean time of residence of a cargo molecule on the membrane, approximately given by $\bar{T} = \bar{T}_f + \bar{T}_d$. The highest efficiency is obtained when $\bar{T}$ is minimal, i.e., for

$$C \sim C_{\text{opt}} \sim m^{-2}. \tag{3.19}$$

In this optimal regime, each molecule spends approximately the same amount of time freely diffusing in the gas and as a part of a growing domain, i.e.

$$\bar{T} \sim \bar{T} \sim \bar{T}_d \sim \bar{T}_f \sim \left(\frac{mN}{D\phi} \right)^{1/2}. \tag{3.20}$$

For a fixed value of the incoming flux ϕ, the mean total residence time of a molecule on the membrane increases therefore with the number of different species as $\bar{T} \sim N^{1/2}$. In the optimal regime, the average density of freely diffusing molecules and the average density of supercritical domains behave as

$$\bar{n}_{\text{opt}} \sim \left(\frac{m\phi N}{D} \right)^{1/2}, \tag{3.21}$$

$$\bar{N}_{d,\text{opt}} \sim \left(\frac{\phi N}{mD} \right)^{1/2}. \tag{3.22}$$

From (3.15) and (3.16), it follows that also the total molecule density $\rho \sim \bar{n} + m\bar{N}_d$ scales as $\rho \sim N^{1/2}$, and is minimal for $C \sim C_{\text{opt}}$.

A consequence of the above relations is that even for low, fixed values of the *total* incoming molecule flux ϕ, the low-density regime, where molecules and domains of different species do not interact significantly, progressively breaks down as the number N of species increases. In the presence of multiple species, according to Eq. 3.2, the flux of molecules of the i-th species across a domain boundary can be approximately written as $\Phi_i \sim D\bar{n}_i$. For very high N, such expression should be modified into $\Phi_i \propto f_N D\bar{n}_i$, with f_N a decreasing function of N, since the crowding of molecules of different species surrounding a domain can be expected to cause an effective decrease in the flux Φ_i at the surface. In that case, a simple modification of the previous phenomenological arguments gives $C_{\mathrm{opt}} \sim f_N m^{-2}$, thus predicting that the optimal effective interaction between homotypic molecules should decrease for very large N.

Rate of supercritical domain production

The rate at which supercritical domains are generated in the non-equilibrium driven stationary state of the lattice model of molecular sorting is assumed, in Eq.(3.9), to be proportional to the square of the free molecule density $\bar{n}$. Here a justification for this assumption is provided, based on a simplified mean-field model of monomer aggregation.

Consider a model of domain formation by means of monomer attachment and detachment, and suppose that there exists a threshold area value A_c above which monomer detachment from domains is not possible, and clusters grow irreversibly. This way, the existence of a critical size in the system is artificially reproduced. Let us call n_A the number density per unit surface of domains of area A, and N_+ the number density per unit surface of domains with $A > A_c$. The incoming flux of monomers is ϕ. The set of mean-field Smoluchowski equations for this model is:

$$\frac{dn_1}{dt} = -2c_1 n_1^2 + b_2 n_2 - n_1 \sum_{A=2}^{A_c} c_A n_A + \sum_{A=2}^{A_c} b_A n_A - c_+ n_1 N_+ + \phi \tag{3.23a}$$

$$\frac{dn_A}{dt} = c_{A-1} n_1 n_{A-1} - b_A n_A - c_A n_1 n_A + b_{A+1} n_{A+1}, \qquad 2 \leq A \leq A_c \tag{3.23b}$$

$$\frac{dN_+}{dt} = c_{A_c} n_1 n_{A_c} \tag{3.23c}$$

where c_A (and, respectively, b_A) are dimensional coefficients representing the attachment (detachment) rates of monomers on (from) domains of area A, and $b_A = 0$ for $A > A_c$. According to reaction rate theory (Sec. 2.4), in two dimensions, the effective reaction rate of two domains is proportional to the sum of their diffusion constants. In the approximation where only monomers can move (extended domains being much slower, as their diffusivity decreases with size as $A^{-3/2}$), the effective aggregation rates c_A become independent of A and proportional to the diffusivity D of a monomer. It is also important to notice that dimers can split with a rate proportional to D/g per molecule, i.e. $b_2 = b/g$. Summing over the areas $2 \leq A \leq A_c$ to obtain an equation for $N_- = \sum_{A=2}^{A_c} n_A$, we find

$$\frac{dN_-}{dt} = c\,n_1^2 - c\,n_1 n_{A_c} - g^{-1} b\,n_2 \tag{3.24a}$$

$$\frac{dN_+}{dt} = c\,n_1 n_{A_c}. \tag{3.24b}$$

The stationary condition $dN_-/dt = 0$ for the subcritical domains implies $dN_+/dt = c\,n_1^2 - g^{-1} b\,n_2$. In order for N_- to be approximately constant with a non-zero production of supercritical domains, the second term must be subdominant already at moderately large values of g. One can then conclude that the net production of supercritical domains N_+ is well approximated by the equation $dN_+/dt \approx c\,n_1^2$. The quantity N_+ corresponds to the number density N_d of supercritical domains used in the main text, thus qualitatively justifying Eq. 3.9.

Entropy production

The ordering effect of the sorting process on the molecule gas can be quantified as follows. Consider the case where molecule injection takes place by the fusion into a membrane system of area A of vesicles carrying a well-mixed cargo of molecules of all of the N distinct molecular species. In the steady state, the same number of cargo molecules is extracted in vesicles containing in average $\delta \mathcal{N}_i = A\,\phi_i\,\delta t$ molecules of only one of the i-th molecular species. Formally treating empty membrane regions of area A_0 as a 0-th molecular species, we define a corresponding flux density ϕ_0. The change in entropy due to the demixing process can then be measured as [61, 101]:

$$\delta S = \sum_{i=0}^{N} \delta \mathcal{N}_i \log \frac{\delta \mathcal{N}_i}{\delta \mathcal{N}}, \tag{3.25}$$

with $\delta\mathcal{N} = \sum_{i=0}^{N} \delta\mathcal{N}_i$, resulting in a simple expression for the average rate of entropy production per unit membrane area:

$$
\begin{aligned}
\frac{1}{A}\frac{\delta S}{\delta t} &= \sum_{i=0}^{N} \phi_i \log \frac{\phi_i}{\phi_0 + \phi} \\
&= -\phi_0 \log\left(1 + \frac{\phi}{\phi_0}\right) - \phi \log\left(1 + \frac{\phi_0}{\phi}\right) - \phi \log N.
\end{aligned}
$$

3.2 Lattice-gas model

Numerical investigations of the molecular sorting process are performed by means of a minimal hard-core lattice-gas model introduced in [7]. The lipid membrane is represented as a two-dimensional square lattice consisting of L^2 sites with periodic boundary conditions. Let η_i be a binary variable denoting the presence or absence of a molecule on site i and let $f(\eta)$ be a function of the state $\eta = \{\eta_i\}$. The evolution of the system is given by a continuous Markov process described by the infinitesimal generator $\mathscr{L} = \mathscr{L}_I + \mathscr{L}_D + \mathscr{L}_E$, with $\mathscr{L}_I$, $\mathscr{L}_D$ and $\mathscr{L}_E$ the generators for molecule insertion, diffusion-aggregation and extraction, respectively. The operator $\mathscr{L}$ is defined as

$$\partial_t \langle f(\eta) \rangle = \langle \mathscr{L}(\eta) \rangle, \tag{3.26}$$

where the average is over the realizations of the process. The allowed moves are the following ones (Fig. 3.2):

i) insertion: molecules from an infinite reservoir arrive and are inserted on empty sites with rate k_I:

$$\mathscr{L}_I f(\eta) = k_I \sum_i (1 - \eta_i)[f(\eta^i) - f(\eta)], \tag{3.27}$$

where η^i denotes the configuration η after the insertion of a molecule on site i;

ii) diffusion and aggregation: molecules can perform diffusive jumps to empty nearest neighboring sites N_i:

$$\mathscr{L}_D f(\eta) = k_D \sum_i \sum_{j \in N_i} \eta_i (1 - \eta_j) g_i(\eta)[f(\eta^{ij}) - f(\eta)], \tag{3.28}$$
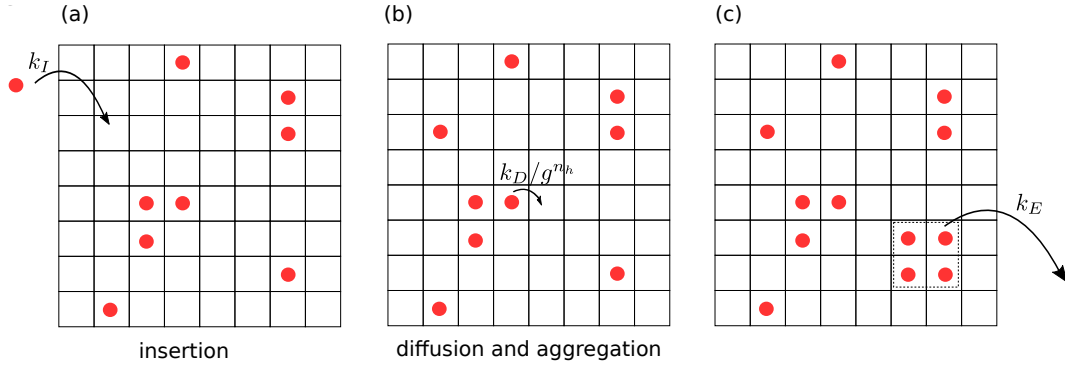
Fig. 3.2 Schematic representation of the lattice-gas implementation of the minimal model of molecular sorting. (a) Molecules are inserted on empty sites with rate k_I. (b) Molecules can jump to an empty neighboring site with rate k_D/g^{n_h}, where g is the intermolecular interaction strength and n_h is the number of homotypic molecules in neighboring sites. (c) Molecules are extracted from the system by the simultaneous removal of all connected clusters of molecules containing a completely filled square of a characteristic size with rate k_E.

where η^{ij} denotes the configuration after the jump has occurred and the diffusion rate k_D is reduced by a factor

$$g_i(\eta) = \prod_{k \in N_i} g^{-\eta_k}, \qquad (3.29)$$

which penalizes the jumps of molecules with nearest neighbours. Domain formation is favoured for values of the interaction strength $g \geq 1$;

iii) extraction: molecules are extracted from the system by simultaneously removing all connected clusters of molecules that contain a completely filled square region V of linear size l with rate extraction k_E:

$$\mathscr{L}_E f(\eta) = k_E \sum_{C \in \mathscr{C}} h\left(\sum_{V \subset C} \prod_{j \in V} \eta_j\right) [f(\eta^C) - f(\eta)], \qquad (3.30)$$

where $\mathscr{C}$ is the connection of connected subsets, $h(x)$ for $x = 0$ and $h(x) = 1$ for $x > 0$, and η^C is the configuration obtained after the extraction (by emptying all the sites in C).

Simulations are performed in the limit $k_E \to \infty$, using Gillespie's algorithm [102].

Critical radius

Assuming that in a quasi-equilibrium condition the molecule gas density outside of a growing domain in a lattice-gas model follows the Gibbs-Thomson relation (Eq. 2.11), where σ is the line tension of the domain, using Eq. 3.3 we get

$$\dot{R} = \frac{A_0 D}{\log(L/R)} \left(R - \frac{\sigma n_\infty}{\bar{n} - n_\infty} \right) \frac{\bar{n} - n_\infty}{R^2}. \tag{3.31}$$

Therefore the critical value of the domain radius is

$$R_c = \frac{\sigma n_\infty}{\bar{n} - n_\infty}. \tag{3.32}$$

This is a non-equilibrium result, in which $\bar{n}$ and n_∞ represent respectively the bulk average density of the molecule gas and the equilibrium density of the gas at the interface with a large flat domain.

These two quantities, together with σ, can be easily estimated at equilibrium in a lattice-gas model. Consider an equilibrium lattice-gas model with a chemical potential μ, and let $\varepsilon > 0$ be the energy gain due to the attractive interaction between two molecules occupying nearest-neighboring sites of the lattice. The energy function of the equilibrium lattice-gas system takes the form

$$E(\eta) = \mu \sum_i \eta_i - \varepsilon \sum_{\langle i,j \rangle} \eta_i \eta_j, \tag{3.33}$$

where $\eta = \{\eta_i\}$ with $\eta_i \in \{0,1\}$ for $i = 1, \ldots, \mathcal{N}$ is a binary configuration representing the presence or absence of molecules on lattice sites. According to the dynamic viewpoint of Ref. 103, the expression

$$R_c \approx \frac{\varepsilon}{z\varepsilon - 2\mu}, \tag{3.34}$$

with z the number of nearest neighbors of a given site, is obtained imposing a local equilibrium condition between the probability of growing and that of shrinking. In a mean-field equilibrium picture, the chemical potential is related to the average total density $\bar{n}_{eq} = e^{-\beta\mu}$ of free molecules in a supersaturated system. At the condensation point $\mu = z\varepsilon/2$, the average total density is equal to the saturation density $n_\infty \approx e^{-\beta z\varepsilon/2}$, which is the molecule density of a gas phase in equilibrium

with a liquid phase (with flat interface). In terms of these quantities, the critical domain radius becomes

$$R_c \approx \frac{\beta \varepsilon}{2 \log(\bar{n}_{\text{eq}}/n_\infty)} \approx \frac{\beta \varepsilon n_\infty}{2 (\bar{n}_{\text{eq}} - n_\infty)} \tag{3.35}$$

close to the condensation point. The expression is formally equivalent to Eq. 3.32 if we identify $\sigma = \beta \varepsilon / 2$. Given two configurations η, η' of the dynamic lattice-gas model, the detailed balance condition implies

$$\frac{W(\eta \to \eta')}{W(\eta' \to \eta)} = \frac{P_{\text{eq}}(\eta')}{P_{\text{eq}}(\eta)}. \tag{3.36}$$

Focusing on the transition $\eta \to \eta'$, in which a dimer fragments into two monomers as a consequence of one of them hopping away, and since $z = 4$ for a square lattice, the previous relation implies $\beta \varepsilon = \log g$, and consequently $n_\infty \approx e^{-2\beta\varepsilon} = g^{-2}$. Moreover, in a lattice gas at equilibrium, the average density $\bar{n}_{\text{eq}}$ of supersaturated gas is fixed by the chemical potential and is independent of the microscopic interaction strength. Therefore, from (3.35) the critical radius is seen to be a monotonically decreasing function of g.

Chapter 4

Phase separation and critical size in molecular sorting

The content of this Chapter is based on the published paper [90].

In the abstract model of the sorting process described in Chapter 3, molecules arriving on a membrane region can laterally diffuse and aggregate into localized domains, whose formation and growth occurs through the typical stages of phase separation: after the initial nucleation stage, in the case of low supersaturation, the growth of domains is mainly governed by the absorption of freely diffusing molecules. One of the main predictions of the classical theory of phase separation is that a *critical size* A_c has to be reached in order for domains to survive and continue to grow irreversibly to larger and larger scales (Sec. 2.4). In the present theory of molecular distillation such domains are extracted once they reach a characteristic size $A_E \gg A_c$, determined by the physical and biomolecular processes that induce membrane bending and vesicle formation. In the presence of a constant flux of incoming molecules, the membrane system selforganizes in a driven non-equilibrium stationary state, which can be seen as a realization in Nature of the classical Szilard's model of droplet formation [104, 105, 62].

Phase separation phenomena are emerging as central drivers of the selforganization of cell structures and the idea that phase separation is an essential step for molecular sorting is increasingly finding support in recent studies (Chapter 2). As advances in live-cell imaging have enabled more accurate observations in real time, a striking

heterogeneity in domain growth kinetics has emerged, and several approaches to unambiguously classify different dynamic populations have been proposed [106–111]. In the experiments, a crucial parameter used to describe the sorting process is the lifetime of a sorting domain. It has been recently shown that the lifetime of a sorting domain is related to the domain stability, which in its turn depends on the number of molecules contained in the domain, and thus on the domain size [112].

It is therefore tempting to relate the existence, in the context of phase separation, of a critical size for domain growth, to the observation that sorting domains on cell membranes can undergo qualitatively different final fates. As a matter of fact, sorting domains are commonly classified in two groups: *productive* domains, if their growth eventually terminates in the nucleation of a vesicle which is ultimately detached from the membrane, and *unproductive* (or abortive) domains which, instead, progressively dismantle and are ultimately dissolved [113, 107, 111]. It seems natural to interpret this distinction in the context of classical nucleation theory, where the fate of a domain results from the balance between bulk stabilization and the propensity to dismantle along the domain boundary, which in its turn is controlled by the value of a characteristic boundary tension (Chapter 2). As a result, circular domains (that minimize the boundary perimeter) are favored, subcritical domains (having size $A < A_c$) have short lifetimes and a low probability of reaching the extraction size A_E, while supercritical domains have a high probability of being ultimately extracted. Here we discuss the implications of this picture in the framework of the phenomenological theory of molecular sorting described in Chapter 3.

In Sec. 4.1 several predictions of the phenomenological theory are verified by extensive numerical simulations of the lattice-gas model introduced in Sec. 3.2. To help in the analysis of experimental data, in Sec. 4.2 we introduce an operational definition of critical size, and its relation to recently introduced methods for the classification of domain formation events into productive and unproductive classes [111] are discussed in Sec. 4.3. The operational definition is used here to compare the predictions of our phenomenological theory of molecular sorting to experiments on the formation of productive and unproductive clathrin-coated pits at the plasma membrane. However, the proposed framework is more general, and we expect that it can turn useful in the interpretation of experiments on molecular sorting at different membrane regions, such as sorting endosomes, or the Golgi complex.

4.1   Scaling laws

One of the main observations of the phenomenological theory of molecular sorting (Sec. 3.1) is that both the average permanence time $\bar{T}$ of sorted molecules on the membrane system and the average molecule density ρ in the steady state are minimal in an intermediate, optimal range of values of the interaction strength g, where the molecular distillation process is most efficient. This is confirmed by numerical simulations of the lattice-gas model (Sec. 3.2). In what follows, areas are measured as numbers of lattice sites, i.e. $A_0 = 1$, and the characteristic area of extracted domains is $A_E = mA_0$ with $m = 10^2$. In every simulation, the system is allowed to relax to the steady state before starting the collection of relevant statistical data. Snapshots of the simulations taken in the steady state show the typical behavior of the system both inside and outside of this optimal range (Fig. 4.1).

(a)   (b)   (c)   (d)   (e)

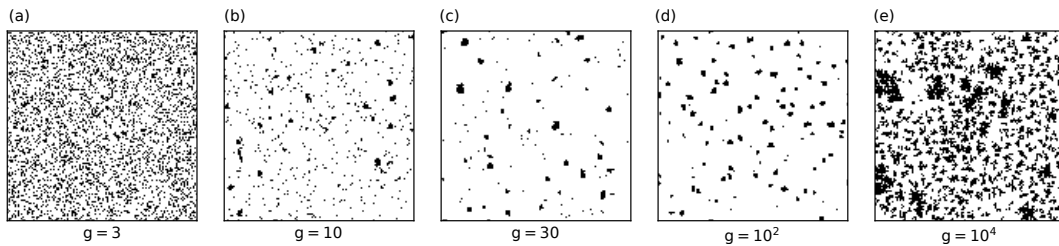g = 3   g = 10   g = 30   g = 10^2   g = 10^4

Fig. 4.1 Snapshots of configurations of the lattice-gas model of molecular sorting for a system of 400^2 sites in the steady state, with incoming flux $\phi/k_D = 10^{-6}$ and increasing values of the interaction strength g (from left to right). In the central panel the interaction strength is close to the optimal value $g_{\text{opt}} = 30$.

For low interaction strength, molecular crowding accompanied by a hectic formation of small short-lived domains is observed (Fig. 4.1(a)). As the interaction strength increases, the density of freely diffusing molecules decreases (Fig. 4.1(b-d)). Consistently with the predictions of the phenomenological theory, the molecular density ρ and residence time $\bar{T}$ are lower in this intermediate range, and reach a minimum in correspondence with the optimal value of the interaction strength g (Ref. [7] and Fig. 4.1(c)). When the interaction strength becomes much larger than its optimal value, the gas of free molecules is strongly depleted, and the system enters into a regime of domain crowding (Fig. 4.1(d)). Here, a large number of sorting domains shares the incoming molecular flux, the growth of each sorting domain is slowed down, and the efficiency of the distillation process is impaired, as

both the molecular density and molecular residence time are much larger than in the optimal region. For very high values of the microscopic interaction strength g, the formation of highly irregular domains of the type predicted by DLA theory [114] is observed (Fig. 4.1(e)). This latter regime is unlikely to correspond to physiological sorting, but could be related to pathological conditions where high intermolecular interaction strength induced by mutations promote the formation of irregular, solid-like aggregates associated to degenerative deseases [2, 32]. Similar behaviors have also been observed in experiments, where overexpression of adaptor proteins responsible for mediating intermolecular interactions leads to the formation of large and irregularly shaped sorting domains [115]. In summary, for varying values of the interaction strength g, our abstract model recapitulates two main phenomenologies. At low and intermediate values of the interaction strength g, the simulated dynamics is characterized by the formation of approximately circular sorting domains via nucleation and coarsening, compatibly with the phenomenology of liquid-liquid phase separation observed in several important biological processes [96, 116, 35]. For very large g instead, domain remodeling is impaired and a DLA phenomenology [114] is recovered, which may possibly describe the features of pathological processes. A precise characterization of the crossover between these two regimes will be the matter of future investigation.

Numerical simulations confirm the validity of the scaling laws $\rho_{\text{opt}} \sim \phi^a$, $\bar{n}_{\text{opt}} \sim \phi^b$ and $\bar{T}_{\text{opt}} \sim \phi^{-c}$, as the numerically obtained values $a = 0.48$, $b = 0.46$ and $c = 0.52$ are in good agreement with the theoretical predictions $a = b = c = 1/2$ [7], that were derived under simplifying assumptions.

In addition to these former results, other predictions of the phenomenological theory can be verified numerically using the microscopic lattice-gas model. The previously exposed phenomenological theory is valid in the regime where supercritical domains are well separated objects, with a well defined value of the average interdomain half distance $\bar{L}$. Since the number of supercritical domains scales as $\bar{N}_d \sim \phi^{1/2}$, and $\pi \bar{L}^2 \bar{N}_d \approx 1$, it is expected that $\bar{L} \sim \phi^{-1/4}$. This scaling law can be verified numerically in the following way. First, the center of mass of each domain is computed. A critical size is determined using the operational definition given in the following Sec. 4.2. Domains with size smaller than the critical size are neglected. The nearest neighbour of each domain is found (Fig. 4.2(a)). Finally, the distances between nearest neighbors and the corresponding statistical measures are computed. The numerical values of the average interdomain half distance $\bar{L}$ obtained
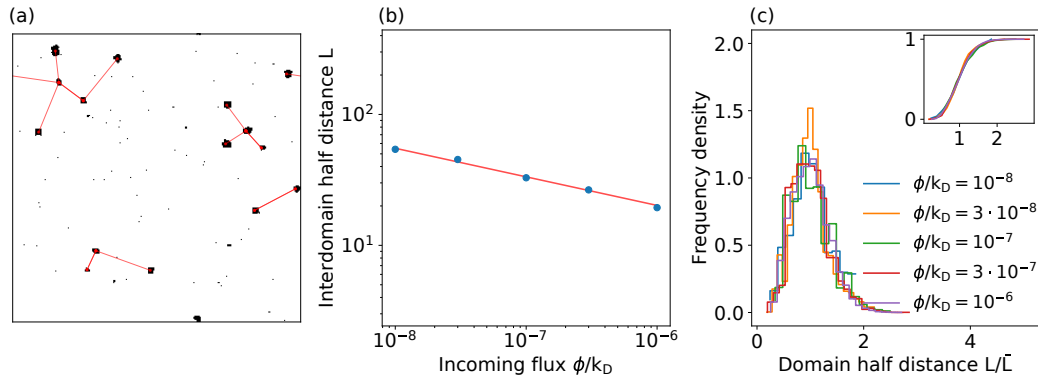
Fig. 4.2 (a) Nearest-neighbor distances between simulated sorting domains are highlighted in red in a snapshot from a simulation performed with incoming flux $\phi/k_D = 10^{-7}$ and interaction strength $g = 10^2$. (b) Scaling of the optimal values of the average interdomain half distance. The red solid line is a fit with the power law ϕ^{-a}, with $a = 0.23$. (c) The frequency density and cumulative frequency distribution (inset) for the rescaled half distances $L/\bar{L}$ for varying values of the incoming flux ϕ/k_D collapse on a single universal frequency distribution.

by this method follow a scaling law $\bar{L} \sim \phi^{-d}$ with $d = 0.23$, close to the theoretically predicted value $d = 1/4$ (Fig. 4.2(b)). When the mean value $\bar{L}$ is used to rescale the interdomain half distances, the corresponding frequency distributions for different values of ϕ collapse on a single universal distribution (Fig. 4.2(c)).

Several results of the phenomenological theory stem from the assumption that the steady-state profile of molecule density around a sorting domain has the logarithmic form (Eq. 3.1), and from the related idea that the membrane region can be divided into "attraction basins" of linear size $\sim L$ pertaining to distinct sorting domains. Given the approximate nature of these hypotheses, it is interesting to check their validity by direct numerical simulations. A convenient way to computationally define this kind of attraction basins is the use of a Voronoi decomposition, which is a partition of the plane into non-overlapping regions according to their proximity to points of a given set [117]. The two-dimensional square lattice used for the numerical simulations was therefore decomposed according to the following procedure. Once all supercritical domains were identified and tracked, for each time frame the center of mass of each domain was computed and the set of these centers was used to partition the lattice area into Voronoi regions (Fig. 4.3(b)). Then, free molecules belonging to each region were identified, and their distance from the domain center of mass computed. A direct validation of the theoretical expression (Eq. 3.1) is computationally very demanding,

as it requires building histograms of distances conditional to the radius R of a given sorting domain. We studied a slightly different quantity, i.e. the average frequency of the distances of free molecules from domains of linear sizes R comprised between the critical radius R_c and the extraction radius R_E:

$$\bar{n}(r) = \int_{R_c}^{R_E} n_R(r) N_{st}(R)\, dR \tag{4.1}$$

for $0 \leq r \leq L$, where the theoretical model describes a density profile characterized by gas depletion in the proximity of the sorting domain. Computing the integral in (4.1) we obtain

$$\bar{n}(r) = K_1 + K_2 \log(r), \tag{4.2}$$

where K_1 and K_2 are functions of the model parameters. If $p(r)\, dr$ is the empirical probability of finding a molecule at a distance comprised between r and $r + dr$ from the center of mass of a domain, then

$$\bar{n}(r) = \frac{p(r)}{2\pi r}. \tag{4.3}$$

The measure of $\bar{n}(r)$ obtained from the numerical simulations by this procedure is in agreement with a fit of the theoretical prediction (Fig. 4.3(a)).

In the phenomenological theory, a central role is played by the dimensionless effective interaction strength C. A convenient expression for C, amenable to empirical estimation, can be obtained by inverting Eq. 3.9 and making use of Eq. 3.12 to get

$$C = \frac{\phi}{mD\bar{n}^2}, \tag{4.4}$$

which is a function of directly measurable quantities, such as the incoming flux ϕ and the bulk gas density $\bar{n}$. The theory predicts that the optimal value $C = C_{opt}$ scales as m^{-h}, with $h = 2$ (cf. Eq. 3.19). Numerical simulations yield the compatible value $h = 1.8$ (Fig. 4.4(a)).

One of the main tenets of the phenomenological theory is the existence of a well-defined critical domain size A_c, arising from the balance between the mixing power of lateral diffusion and the tendency of sorted molecules to aggregate. In the lattice-gas model, the tendency to aggregation is controlled by the microscopic parameter g, while in the phenomenological theory, an analogous role is played by the effective interaction strength C. The operational definition provided by
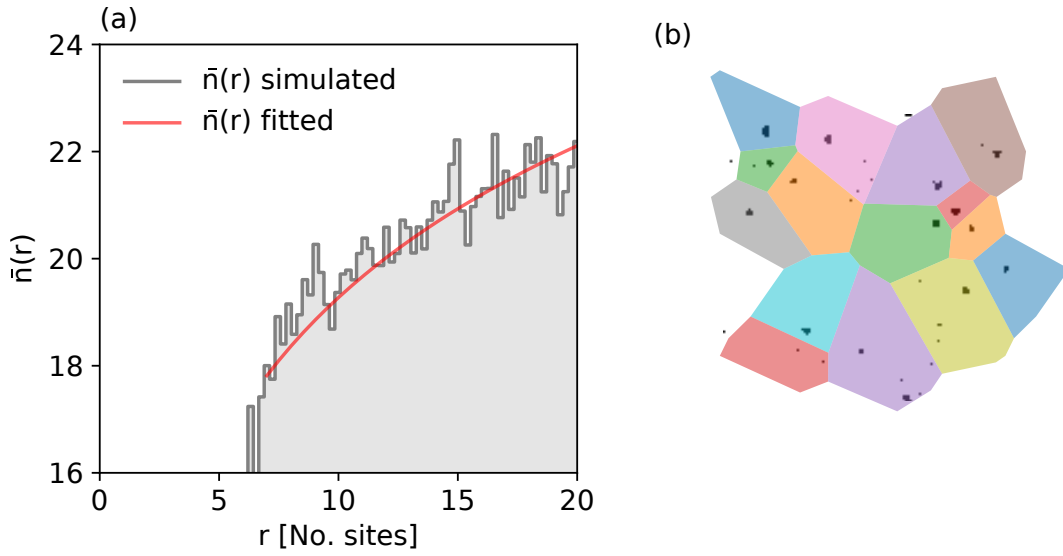
Fig. 4.3 (a) Average density profile $\bar{n}(r)$ of the gas of free molecules at a distance r from the center of supercritical domains, obtained from the simulations, and fitted with the theoretical prediction Eq. 4.2 ($\phi/k_D = 10^{-7}$, $g = 10^2$). (b) Voronoi decomposition obtained from a set of simulated supercritical sorting domains (4.2).

Eq. 4.4 allows to determine C from the simulated molecule density $\bar{n}$ as a function of model parameters (Fig. 4.4(b)). Accordingly with its interpretation as an effective interaction strength, C is observed to be a non linear, monotonically increasing function of the microscopic parameter g.

4.2   Operational definition of critical size

The critical domain size $A_c = \pi R_c^2$ is a central control parameter of the molecular distillation process, but there is no simple analytical expression for it in the framework of the phenomenological theory. Explicit approximate expressions for the critical size can be obtained using classical metastability analysis in quasi-equilibrium lattice-gas models (Sec. 3.2). Such an analysis predicts that A_c is a monotonically decreasing function of the *microscopic* interaction strength between sorted molecules, which, however, is not practically measurable. For this reason, here we provide an operational definition of critical size that can be more directly related to the analysis of experimental observations.
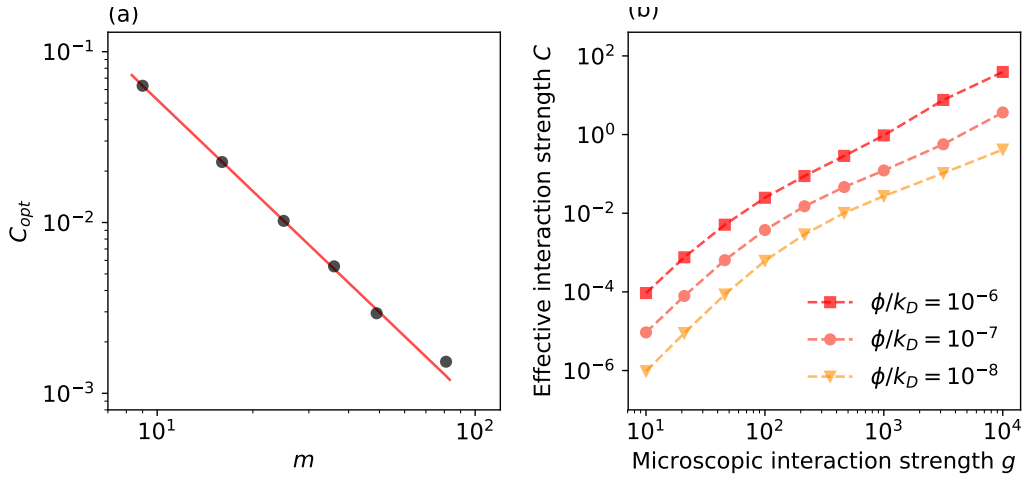
Fig. 4.4 (a) Optimal effective interaction strength C_{opt} as a function of $m = A_E/A_0$, at fixed $\phi/k_D = 10^{-6}$. The red solid line is a fit with the power law m^{-h}, with $h = 1.8$. (b) Effective interaction strength C as a function of the microscopic interaction strength g, for different values of the incoming flux.

In experimental studies of molecular sorting, domain "trajectories" have been observed to fall into two classes, depending on their fate [113, 107, 111]: *productive* trajectories, where the domain is finally extracted as a part of a lipid vesicle, and *unproductive* trajectories, where the domain progressively dismantles and is ultimately dissolved. It is worth observing here that these are properties of the domain *history*, and not of its state at a given instant. However, for simplicity, we will define in what follows as *productive* or *unproductive* domains, those that belong to productive or unproductive trajectories, respectively. In our lattice-gas model, productive and unproductive domains can be directly distinguished by tracking their evolution in time, and checking whether their trajectory ends up with an extraction event, or not (Fig. 4.5). The classification into productive and unproductive trajectories can be used to provide a natural, operational definition of critical size, applicable to the analysis of actual experimental data. Let us define the 'operational' critical size as the value $\mathscr{A}_{\text{c}}$ such that a domain of size $\mathscr{A}_{\text{c}}$ has 50% probability of being productive:

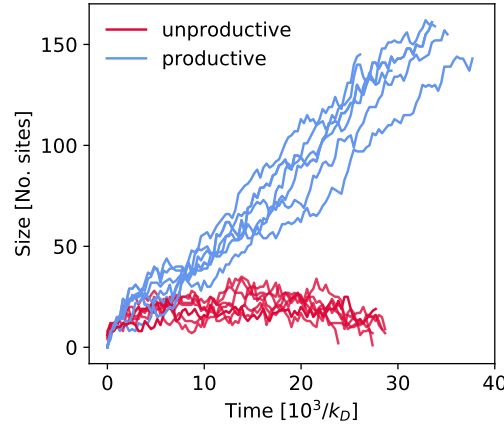$$P(\text{prod.}|\mathscr{A}_{\text{c}}) = \frac{1}{2}, \tag{4.5}$$

Fig. 4.5 Time evolution of the size of productive (blue) and unproductive (red) sorting domains, from numerical simulation of the lattice-gas model ($\phi/k_D = 10^{-6}$, $g = 20$).

(similar definitions have been adopted in previous works, see e.g. Ref. 118). In terms of (joint) probability density functions (pdf's), Eq. 4.5 is equivalent to

$$p(\mathscr{A}_c, \text{prod.}) = p(\mathscr{A}_c, \text{unprod.}), \qquad (4.6)$$

i.e., the critical size is found at the intersection of the joint pdf's of, respectively, productive and unproductive domain sizes. Under a few additional hypotheses (discussed at the end of this Section), Eq. 4.5 implies

$$P(\text{prod.}|A) \geq \frac{1}{2} \quad \text{for all} \quad A \geq \mathscr{A}_c \qquad (4.7)$$

consistently with the phenomenological picture, where smaller domains decay with high probability, while, once a domain exceeds the critical size, the probability that it will continue to grow up to the extraction size is larger than the probability that it will disappear. In terms of the joint pdf's of, respectively, productive and unproductive domains, Eq. 4.7 is in its turn equivalent to the condition that

$$p(A, \text{prod.}) \geq p(A, \text{unprod.}) \quad \text{for all} \quad A \geq \mathscr{A}_c. \qquad (4.8)$$

Either (4.7) or (4.8) can be conveniently applied to the analysis of empirical data, which are given as integer or floating-point numbers of finite precision. The critical size $\mathscr{A}_c$ can thus be estimated either from conditional frequencies (using Eq. 4.7) or from frequency histograms of domain sizes (using Eq. 4.8), as long as productive and
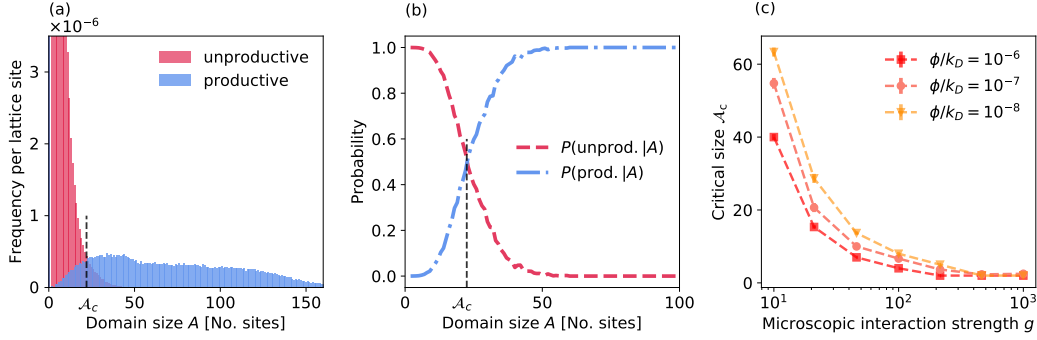
Fig. 4.6 (a) Empirical histograms of domain sizes for productive (rightmost histogram, blue) and unproductive (leftmost histogram, red) domains obtained from numerical simulations of the lattice-gas model ($\phi/k_D = 10^{-7}$, $g = 20$). (b) Probability of a domain being productive or unproductive, conditioned by its size A. The vertical dashed lines mark the position of the critical size $\mathscr{A}_c$, that can be found, according to (4.8), where the frequency of productive domains surpasses the frequency of unproductive domains (a), or equivalently, according to (4.7), where the conditional probability of a domain of size A being productive exceeds $1/2$ (b). (c) Critical size $\mathscr{A}_c$ as a function of the interaction strength g for different values of the incoming flux ϕ/k_D.

unproductive domains can be effectively discriminated. As an example, in Fig. 4.6(a), $\mathscr{A}_c$ is found at the approximate intersection of the (joint) frequency histograms of, respectively, productive and unproductive domains. The existence of this intersection appears to be guaranteed by the fact that $p(A, \text{unprod.})$ is a decreasing function of A, while $p(A, \text{prod.})$ is initially increasing. Fig. 4.6(b) shows that the probability of a domain being productive increases with its size, while the complementary probability of being unproductive decreases. The above procedure allows to compute $\mathscr{A}_c$ from numerical simulations for different values of model parameters. The critical size $\mathscr{A}_c$ is thus found to be a decreasing function of both the incoming molecule flux ϕ, and of the microscopic interaction strength g (Fig. 4.6(c)), consistently with the prediction for the critical size derived in the context of mestability in quasi-equilibrium lattice-gas models (see Sec. 3.2). In the non-equilibrium case of molecular sorting however the gas density $\bar{n}$, being sustained by the constant molecular influx ϕ, decreases slower than the equilibrium gas density n_∞ .

Having at our disposal an operational definition of critical size, we are now in a position to check numerically the validity of theoretical predictions about the shape of the domain size distribution. The theory predicts functionally different forms for the number densities for the size of, respectively, subcritical and supercritical domains. In the subcritical region, transient domains continuously form and dissolve.
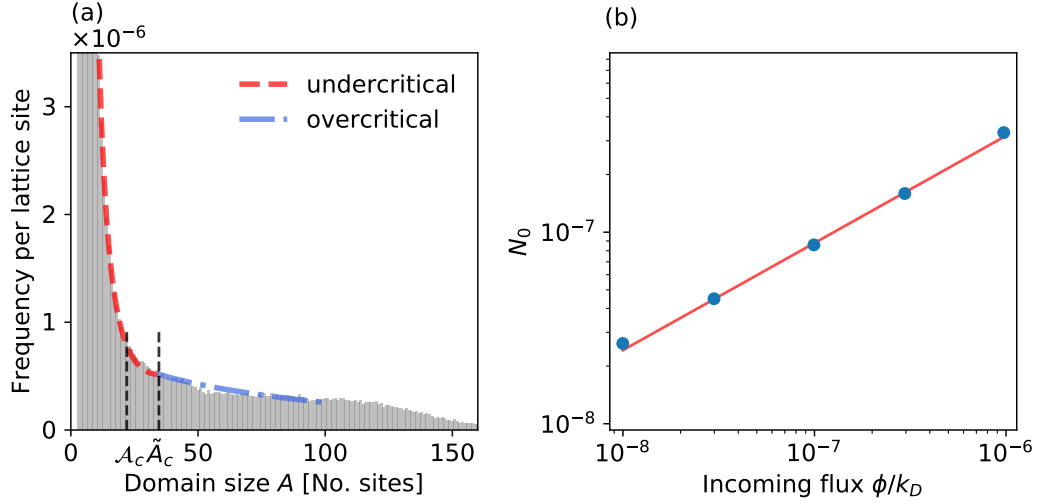
Fig. 4.7 (a) Full histogram of all domain sizes ($\phi/k_D = 10^{-7}$, $g = 20$). The lines are fits with Eq. 4.9 (red dashed line) for $A < \tilde{A}_c$, and with Eq. 4.10 (blue dash-dotted line) for $\tilde{A}_c < A < A_E$. The $A > A_E$ tail depends on the details of the extraction mechanism and is therefore non universal. (b) Numerical estimate of the prefactor N_0 appearing in Eq. 4.10, as a function of the incoming flux ϕ/k_D, in the optimal region. The red solid line is a fit with the power-law ϕ^f with $f = 0.54$.

This quasi-equilibrium state is approximately described by Eq. 2.7, derived in the context of the classical nucleation theory (Sec. 2.4), which predicts that the stationary number density for domains of size $A < A_c$ is

$$N_{st}^{sub}(A) = N_0^{sub} e^{\lambda \left(A^{1/2} - A_c^{1/2}\right)^2},\tag{4.9}$$

where λ is a constant, which is expected to be proportional to the interaction strength between sorted molecules.

For $A > A_c$, according to Eq. 3.5, the shape of the number density is instead of the logarithmic type:

$$N_{st}(A) = N_0 \log \frac{A_L}{A},\tag{4.10}$$

with $N_0 \sim \phi^{1/2}$. By fitting the full histogram of all domain sizes with Eq. 4.9 for small A and with Eq. 4.10 for large A, and by imposing the continuity condition

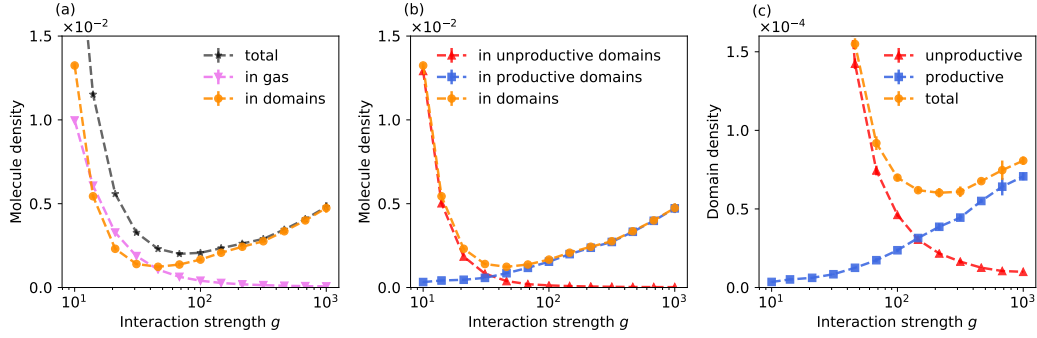$$N_0^{sub} = N_0 \log \frac{A_L}{A_c}\tag{4.11}$$

Fig. 4.8 (a) The number of free molecules per unit area decreases for increasing interaction strength g (magenta triangles), while the number of molecules found inside of sorting domains has an increasing trend at large g (orange dots). As a consequence, the total number of molecules per unit area (black stars) has a minimum, which marks the position of the optimal sorting regime [7]. (b) In its turn, the number of molecules inside of sorting domains (orange dots) is a non-monotonic function of the interaction strength g. This can be understood as follows. The number of molecules inside of unproductive domains (red triangles) decreases with increasing interaction strength, while the number of molecules inside of productive domains (blue squares) increases. As a consequence, the total number of molecules found inside of sorting domains of any of the two types (orange dots) has a minimum close to the optimal sorting regime. (c) Similarly, the number of unproductive domains per unit area (red triangles) decreases with the interaction strength, whereas the number of productive domains (blue squares) increases. As a consequence, the total number of sorting domains of the two types (orange dots) has a minimum for intermediate interaction strength, close to the optimal sorting regime. Simulations performed with $\phi/k_D = 10^{-8}$. The number of both productive and unproductive domains increase with increasing ϕ (not shown here).

one obtains an estimate $\tilde{A}_c$ of the critical size A_c in the framework of classical nucleation theory (see Fig. 4.7(a)). The thus obtained value $\tilde{A}_c$ is of the same order as the previously introduced value $\mathscr{A}_c$, the difference being due to the presence of a small tail of unproductive domains with $A > \mathscr{A}_c$ (see Fig. 4.6(a)). The definition of $\mathscr{A}_c$ has a clear probabilistic interpretation and is independent of phenomenological assumptions about the underlying process of domain formation. On the other hand, the estimate $\tilde{A}_c$ by the above empirical fitting procedure can be used when it is not possible to discriminate between productive and unproductive domains.

A numerical estimate of the prefactor N_0 for different values of the incoming molecule flux ϕ gives $N_0 \sim \phi^f$ with $f = 0.54$, in reasonably good agreement with the theoretical value $f = 1/2$ (Fig. 4.7(b)).
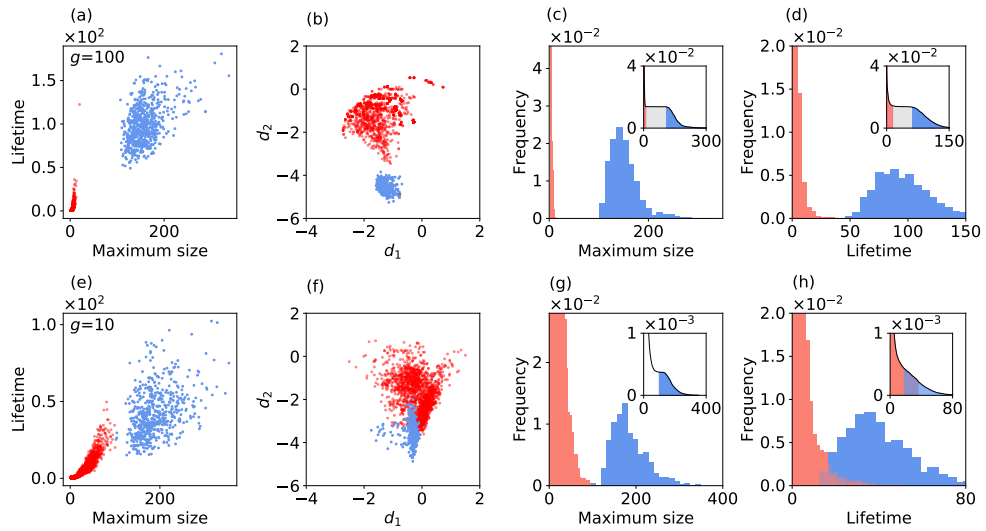
Fig. 4.9 Statistical properties of productive (blue (light gray, rightmost histograms)) and unproductive (red (dark gray, leftmost histograms)) domains for incoming flux $\phi/k_D = 10^{-6}$ and interaction strength $g = 10^2$ (a-d, $5 \cdot 10^4$ domain trajectories) and $g = 10^1$ (e-h, $1.5 \cdot 10^6$ domain trajectories), collected over a $3 \cdot 10^6/k_D$ time interval. Simulated trajectories were classified into productive and unproductive depending on whether they ended up in an extraction event, or not. (a, b, e, f) Scatter plots of domain lifetimes vs. maximum sizes (a, e) and of DASC indicators d_1, d_2 (b, f). (c, d, g, h) frequency distributions of maximum sizes and lifetimes. Insets: complementary cumulative frequency distributions. Domain sizes are given as number of occupied lattice sites, lifetimes are measured in units of $10^3/k_D$.

In addition to these former results, other predictions of the phenomenological theory can be verified numerically using the microscopic lattice-gas model. The previously exposed phenomenological theory is valid in the regime where supercritical domains are well separated objects, with a well defined value of the average interdomain half distance $\bar{L}$. Since the number of supercritical domains scales as $\bar{N}_d \sim \phi^{1/2}$, and $\pi \bar{L}^2 \bar{N}_d \approx 1$, it is expected that $\bar{L} \sim \phi^{-1/4}$. This scaling law can be verified numerically in the following way. First, the center of mass of each domain is computed. A critical size is determined using the operational definition given in the following Sec. 4.2. Domains with size smaller than the critical size are neglected. The nearest neighbour of each domain is found (Fig. 4.2(a)). Finally, the distances between nearest neighbors and the corresponding statistical measures are computed. The numerical values of the average interdomain half distance $\bar{L}$ obtained by this method follow a scaling law $\bar{L} \sim \phi^{-d}$ with $d = 0.23$, close to the theoretically predicted value $d = 1/4$ (Fig. 4.2(b)). When the mean value $\bar{L}$ is used to rescale the

interdomain half distances, the corresponding frequency distributions for different values of ϕ collapse on a single universal distribution (Fig. 4.2(c)).

The systematic discrimination of productive and unproductive domains allows to unravel additional aspects of the phenomenology. Optimal sorting takes place when the total number of molecules in the system is minimal [7] (Fig. 4.8(a)). In a neighborhood of this optimal value, one observes also a minimum in the number of molecules contained in the domains (Fig. 4.8(b)), and in the number of domains itself (Fig. 4.8(c)). This is a somehow paradoxical effect, since at first sight, one would expect that a larger number of sorting domains could increase the speed of the sorting process. Instead, sorting turns out to be most efficient precisely when the number of sorting domains is close to a minimum. As a matter of fact, when the interaction strength increases, the number of molecules in unproductive domains decreases, while the number of those in productive domains increases. As a consequence, their sum, i.e. the number of molecules in any of the two types of domains, has a minimum (Fig. 4.8(b)). A similar argument applies directly to the total numbers of productive and unproductive domains: the number of unproductive domains decreases when the interaction strength increases, while the number of productive domains increases, as predicted by Eq. 3.16 [1]. This leads to the appearance of an intermediate minimum in the total number of domains (Fig. 4.8(c)). The emerging picture is that the efficiency of the sorting process is not favored by a proliferation in the number of sorting domains: in that case, the flux of incoming molecules has to be shared among a larger number of domains, and the growth rate of individual domains is slowed down. A balance has therefore to be struck between two competing requirements: the interaction strength should be large enough to allow for easy nucleation of new sorting domains, but small enough to avoid their unnecessary proliferation.

These theoretical predictions are compatible with former experimental work where the strength of interaction between transferrin receptors on cell plasmamembranes was experimentally controlled, and higher interaction strength was shown to induce higher rates of generation of productive sorting domains, and lower numbers of unproductive events [119].

---

[1]Recalling also that the macroscopic interaction strength C is a monotonically increasing function of the microscopic parameter g (Fig. 4.4(b)).

Alternative definitions of critical size

Let us define the empirical critical size as the value $\mathscr{A}_c$ such that

$$P(\text{prod.}|\mathscr{A}_c) = \frac{1}{2}. \tag{4.12}$$

This value is well defined if $P(\text{prod.}|A)$ is a continuous function which tends to 0 for $A \to 0$ and to 1 for $A \to \infty$. Equivalently, $\mathscr{A}_c$ can be defined as the solution of

$$p(A, \text{prod.}) = p(A, \text{unprod.}) \tag{4.13}$$

since (4.13) can be rewritten as

$$p(A, \text{prod.}) = p(A) - p(A, \text{prod.}) \tag{4.14}$$

yielding

$$P(\text{prod.}|A) = \frac{p(A, \text{prod.})}{p(A)} = \frac{1}{2}. \tag{4.15}$$

If $p(A, \text{unprod.})$ is a decreasing function of A, and $p(A, \text{prod.})$ is an increasing function of A in a right neighborhood of 0 (as the simulations suggest, see e.g. Fig. 4.6), one can easily show that $P(\text{prod.}|A_1 \geqslant A \geqslant \mathscr{A}_c)$ is a non decreasing function of A_1 by directly computing its derivative with respect to that variable. Then, for all $A_1 \geqslant \mathscr{A}_c$ one has:

$$P(\text{prod.}|A \geqslant \mathscr{A}_c) \geqslant P(\text{prod.}|A_1 \geqslant A \geqslant \mathscr{A}_c) \geqslant P(\text{prod.}|A = \mathscr{A}_c) = \frac{1}{2}. \tag{4.16}$$

4.3 Interpretation of experimental data

The correct classification of productive/unproductive trajectories in data obtained from living cell experiments is a challenging process. Several approaches have been adopted. Productive trajectories can be singled out by detecting bursts in the concentration of specific molecules involved in the process of vesicle detachment, such as dynamin [120, 109, 113]. Other approaches rely on the measure of extremal properties of domain trajectories, such as the maximum size reached by domains, or their lifetime [113, 106, 107, 121, 7], which are expected to be less dependent on the small-scale details of the stochastic process. More recently, a new classification

method based on a "disassembly asymmetry score" (DASC) [111] has been proposed. In this context, productive and unproductive trajectories are discriminated by clustering the values of a set of statistical indicators that compare properties of the backward and forward histories of the domains [111]. The effectiveness of some of these approaches can be tested on numerical simulations of the lattice-gas model discussed in the previous Sections, where the productive vs. unproductive classification can be performed exactly. The first two columns of Fig. 4.9 show scatter plots of maximum size vs. lifetime, and of the DASC indicators d_1, d_2 [111], for $g = 10^2$ (a-d) and $g = 10^1$ (e-h). Different colors are used for productive (blue) and unproductive (red) trajectories. For $g = 10^2$ the two populations are clearly separated, and can be easily discriminated automatically using standard clustering methods. For $g = 10^1$ instead the representative points of the two populations start to overlap, and clustering methods are likely to return a certain number of erroneously classified points. For $g = 10^2$ the existence of two distinct populations of domain trajectories is reflected in the bimodal shape of the frequency distributions of maximum sizes and lifetimes (Fig. 4.9(c, d)). This clear separation corresponds to a distinct plateau in the (complementary) cumulative frequency distribution (insets). For $g = 10^1$ instead (Fig. 4.9(g, h)), the frequency distributions of the two populations start to overlap and the bimodal character of the two frequency distributions tends to disappear. The loss of discriminating power takes place approximately for values of the interaction strength such that the critical size $\mathscr{A}_c$ becomes of the order of the extraction size A_E (cf. Fig. 4.6(b)).

Interestingly, the model predictions for the frequency distributions of the maximum sizes and lifetimes of sorting domains are similar to those resulting from experimental observations. In particular, the maximum size and lifetime distributions for unproductive domains show a rapid monotonic decay, while the corresponding distributions for productive domains show a distinct maximum and a slower decaying tail (Fig. 4.9(c, d, g, h)). Both of these features have been observed in experiments of endocytic sorting [107, 111], where productive and unproductive domains correspond to clathrin-coated pits (CCPs) and abortive coats (ACs), respectively. (A third population of outlier traces (OTs) [111], characterized by short lifetimes and large sizes, likely correspond to cytoplasm-originated events [121] and are not observed in the simulations.) We looked for model parameters providing the best fit of simulated frequency distributions with data from Fig. 2B,C of Ref. 111, where productive and unproductive domains were classified using DASC. By a single fit of the two

parameters of the model and of two rescaling factors for the time and length scales, good agreement between simulation and experimental data was found for both the lifetime and maximum size distributions, simultaneously for both productive and unproductive domains (Fig. 4.10). The frequency histograms obtained from the exact classification of simulated productive and unproductive domains (Fig. 4.10(a, b)) was compared with the frequency histograms obtained with the same model parameters, where however simulated domains were classified by the DASC method, yielding similar results (Fig. 4.10(c, d)).
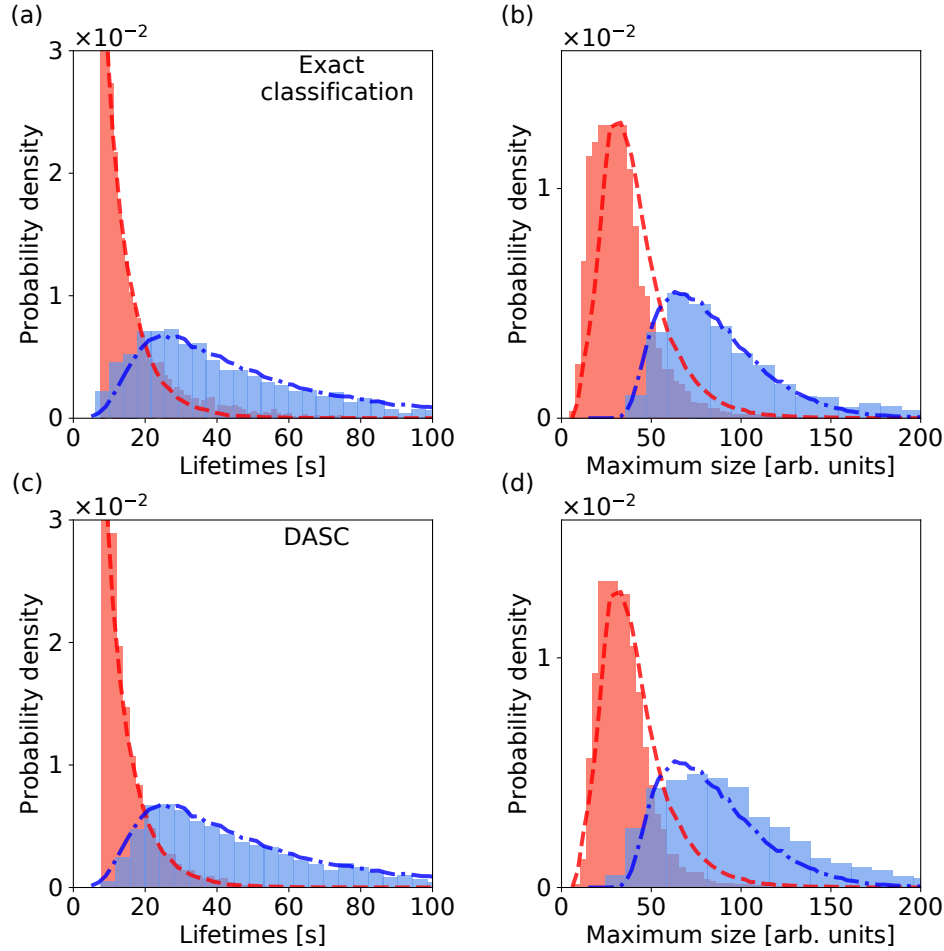
Fig. 4.10 Comparison between the experimental distributions of lifetimes (a, c) and maximum sizes (b, d) of unproductive (red dashed lines) and productive (blue dash-dotted lines) domains from Ref. 111, Fig. 2B,C (kindly shared by Dr. Xinxin Wang), and corresponding distributions obtained from simulations of the lattice-gas model (leftmost red and rightmost blue histograms, respectively) with fitted values of the model parameters ($g = 6.5$, $\phi/k_D = 10^{-6}$) and fitted rescaling factors for lifetime and domain size units ($k_D = 715\,\mathrm{s}^{-1}$, 1 lattice site = 0.3 a.u.). Lower cutoffs on lifetime and maximum size approximately equal to the values reported in the experimental data were used. In the experiments, productive and unproductive domains were classified by DASC. In the analysis of simulated data (histograms), use was made of both the exact classification obtained directly from the simulations (a, b), and a posteriori use of DASC on the numerically generated domains (c, d), obtaining similar results.

Chapter 5

Sorting of multiple molecular species

The content of this Chapter is based on the paper under review [91].

In the past, experimental investigations have been mainly focused on the process of sorting of single molecular species, such as transferrin receptors or low density lipoproteins [7, 5]. More recently, advances in imaging technologies have made it possible to elucidate aspects of the simultaneous distillation of distinct molecular species, and to directly observe their localization in distinct, separate sorting domains [8–11]. The demixing of distinct molecular species subject to attractive homotypic interactions at equilibrium has been also predicted by statistical physics arguments [122, 123].

Here, we investigate the non-equilibrium scenario of sorting N molecular species by combining theoretical arguments and lattice-gas numerical simulations in Sec. 5.1, while in Sec. 5.2 the case where each species is characterized by different microscopic interaction strengths is considered.

5.1   Sorting multiple molecular species

A rough indication about the number of different species that can be sorted in parallel without significantly interfering with each other in the optimal sorting regime can be obtained by noticing that in the low-density regime, the interdomain half distance L has to be much larger than the extraction size $A_E^{1/2}$. Observing that $\pi L^2 \bar{N}_d \sim 1$, and
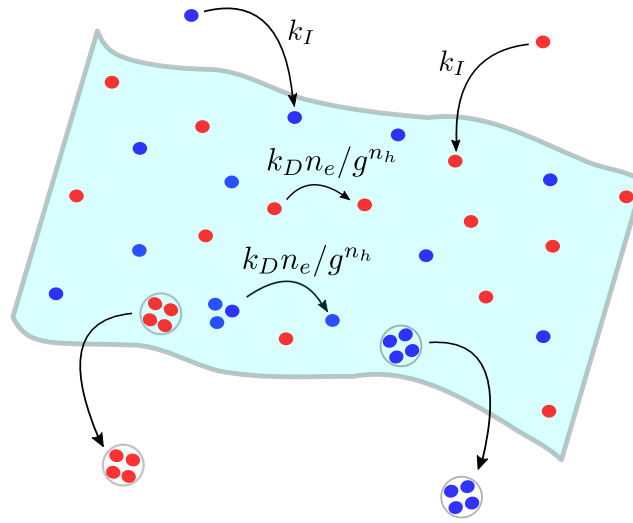
Fig. 5.1 Schematic representation of the hard-core lattice-gas model used to investigate the parallel sorting of multiple molecular species.

using Eq. 3.22, this condition translates into

$$L \sim \bar{N}_d^{-1/2} \sim \left(\frac{Dm}{N\phi}\right)^{1/4} \gg A_E^{1/2} \tag{5.1}$$

giving

$$N \ll \frac{D}{A_E A_0 \phi} \tag{5.2}$$

Using the realistic orders of magnitude $D \sim 10^{-3}$ μm²/s, $A_E \sim 10^{-1}$ μm², $A_0\phi \sim 10^{-5}$ s^{-1} inferred in a previous study [7] from experimental data on the endocytic sorting of low-density lipoproteins, one obtains $N \ll 10^3$. Therefore, the analytical estimate (5.2), based on our phenomenological theory of sorting (Sec. 3.1), suggests that a large number of protein cargos of different species may in principle be distilled in parallel on a cell membrane. It is worth observing, however, that Eq. 5.2 can provide only a qualitative indication about the breakdown of the low-density regime, as it was derived neglecting the contribution of complicated logarithmic prefactors [7]. The onset of the regime of molecular crowding is therefore more precisely investigated by means of numerical simulations of a generalized version of the hard-core lattice-gas model previously introduced (Sec. 3.2). A schematic representation of the stochastic processes taking place in the model is shown in Fig. 5.1.

Molecules of multiple species are distributed on a two-dimensional regular lattice with periodic boundary conditions. Each lattice site can host a single molecule at most, but there is no limit to the number of species that can populate the lattice. The current state of the system is described by a multivariate configuration where 0 marks an empty lattice site, while a number $i \in \{1, \dots, N\}$ marks a molecule of the i-th species residing on a given site. The state of the system evolves according to a continuous-time Markov chain consisting of the following three processes:

i) molecules, whose species are chosen randomly from a set of N species in such a way that the overall flux has the assigned value ϕ, are inserted into empty sites with rate k_I;

ii) molecules jump towards empty neighboring sites with rate $k_D n_e / g_i^{n_h}$, where n_e is the number of empty neighboring sites, g_i is the intermolecular interaction strength between molecules belonging to the i-th species, with $i = 1, ..., N$, and n_h is the number of homotypic molecules in neighboring sites;

iii) the molecules in a connected domain of homotypic molecules are extracted when the domain reaches the extraction size m.

The extraction mechanism adopted here is a slightly simplified version of the one described in Sec. 3.2 and used in Chapter 4, where the molecules of a connected domain were extracted only when the domain grew to the point of containing a molecule-filled square of given size. In the simulations, $A_0 = 1$, i.e. areas are measured as multiples of the elementary lattice area A_0, and the realistic value $m = 25$ is used. This way, the relevant microscopic parameters describing the process are the intermolecular interaction strength g and the ratio k_I/k_D. For low values of the molecule density ρ, such as those experimentally measured in Ref. [7], the molecule flux $\phi = k_I(1-\rho)$ is approximately equal to the insertion rate k_I. Compatibly with experimental observations [7], simulations were performed with $\phi/k_D = 10^{-5}$.

The results of numerical simulations of the model on a 100×100 square lattice are displayed in Fig. 5.2. Consistently with previous results (Ref. [7] and Sec. 4.2), Fig. 5.2a shows that, in an intermediate range of values of the microscopic interaction strength g, the stationary density of molecules ρ exhibits a minimum, corresponding to an optimal sorting regime. Increasing the number N of sorted species, the optimal region moves towards larger values of g, indicating that efficient molecular sorting
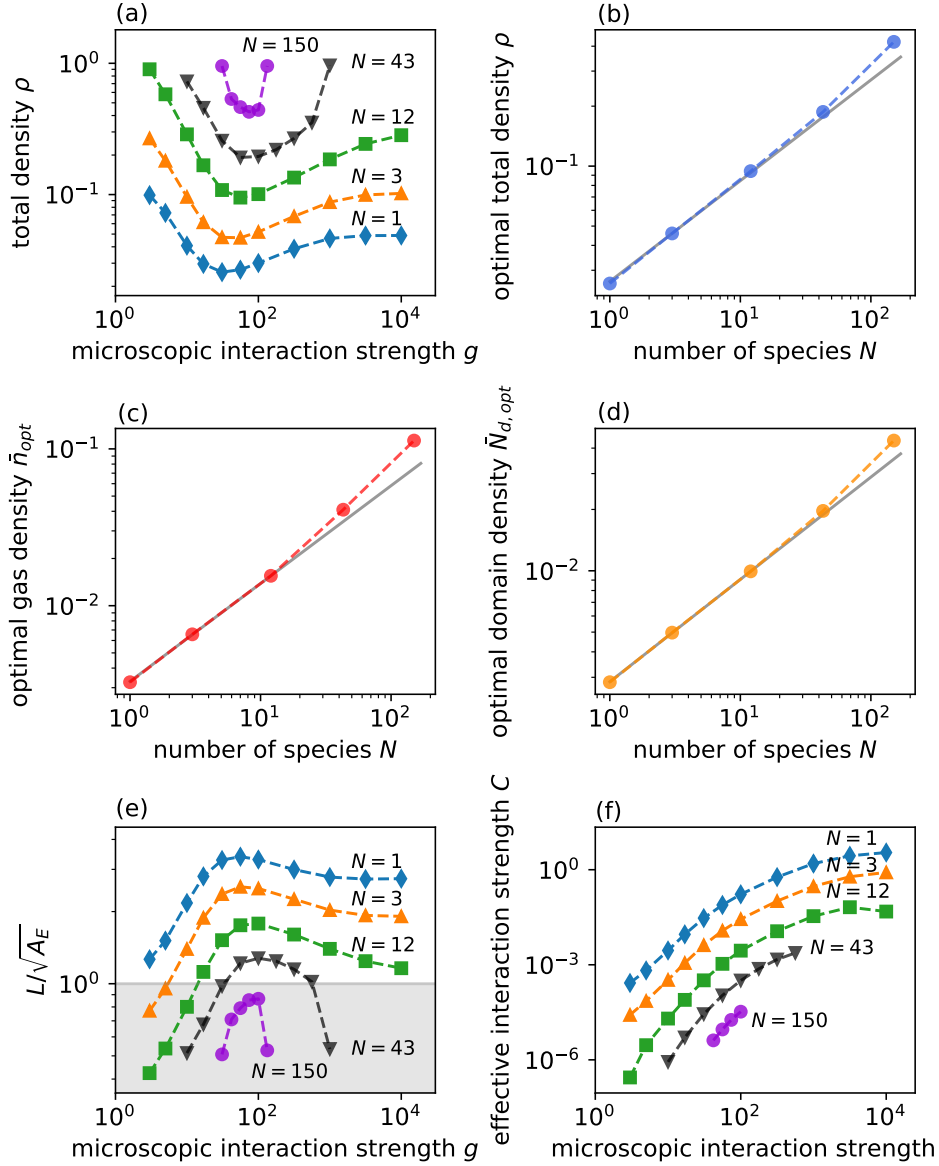
Fig. 5.2 (a) Total molecule density as a function of the microscopic interaction strength g, for increasing values of the number of species N. For large N, sorting remains possible only in a restricted interval of values of g. (b,c,d) Total molecule density, density of freely diffusing molecules, and density of domains, respectively, as functions of the number N of sorted molecular species, in the optimal sorting regime. The solid lines are fitted with $\rho \sim N^a$, $\bar{n} \sim N^b$, and $\bar{N}_d \sim N^c$ with $a = 0.51$, $b = 0.62$, $c = 0.50$. These scaling relations are verified for $N < 20$, whereas, in qualitative agreement with the theory, increasing deviations are observed for larger N. (e) Ratio of the interdomain half-distance L to the extraction size $A_E^{1/2}$ as a function of the microscopic interaction strength g. The parameter region where the system becomes overcrowded and sorting is impaired is marked in light gray. (f) Effective interaction strength C as a function of the microscopic interaction strength g for different values of the number of molecular species N. Simulations were performed at fixed incoming molecule flux $\phi/k_D = 10^{-5}$ on a square lattice.
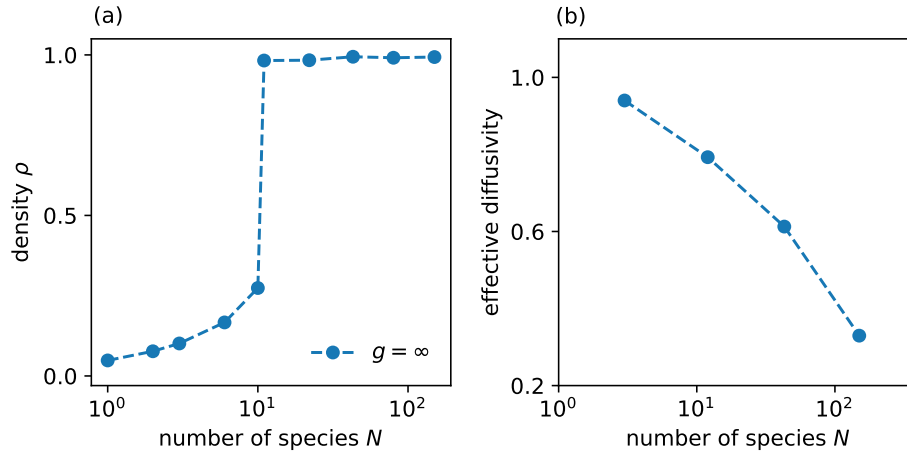
Fig. 5.3 (a) Total molecule density ρ as a function of the number of species N in the limit case $g = \infty$. Above a critical value ~ 10 of the number of species, the density becomes ~ 1, signaling that most lattice sites are filled and that the distillation process has come to a stop due to molecule overcrowding. (b) The effective diffusivity of a test particle decreases as the number of species N increases, signaling that molecular mobility is strongly reduced. Simulations were performed at fixed incoming molecule flux $\phi/k_D = 10^{-5}$ on a square lattice.

becomes more and more difficult to obtain as the number of sorted species increases. For very large N, molecular sorting can take place only in a restricted interval of values of g, as the system tends to freeze into an overcrowded state for both lower and higher values of g. The existence of a maximum value g_N such that sorting at $g > g_N$ becomes impossible for large N due to crowding effects can also be checked by looking at the behavior of the stationary molecular density $\rho_{g=\infty}$ as a function of N (Fig. 5.3(a)): this density rapidly transitions to values of order 1 (high molecular crowding) for $N \sim 10$, signaling that sorting is strongly impaired at high g for $N \gtrsim 10$.

When $N \sim 10$, sorting is still possible around the optimal region, which however tends to shrink progressively with increasing N, since the optimal molecular density $\rho_{\rm opt}$ grows with N, in quantitative agreement with the scaling law $\rho_{\rm opt} \sim N^{1/2}$ predicted by the phenomenological theory (Fig. 5.2(b)). An analogous behaviour is observed for the optimal density of freely diffusing molecules $\bar{n}_{\rm opt} \sim N^{1/2}$ and the optimal density of domains $\bar{N}_{\rm d,opt} \sim N^{1/2}$ (Fig. 5.2(c,d)). A summary of these scaling laws for the optimal regime is reported in Table 5.1.

Scaling law	Theoretical	Numerical
$\rho_{\text{opt}} = \phi \, \bar{T}_{\text{opt}} \sim N^a$	$a = 1/2$	$a = 0.51$
$\bar{n}_{\text{opt}} \sim N^b$	$b = 1/2$	$b = 0.62$
$\bar{N}_{d,\text{opt}} \sim N^c$	$c = 1/2$	$c = 0.50$

Table 5.1 Scaling laws for the average total molecule density ρ, the average density of freely diffusing molecules $\bar{n}$, the average density of domains $\bar{N}_d$ in the optimal regime as a function of the number of molecular species N are reported in the first column. The corresponding theoretically predicted and numerically computed exponents are reported in the second and third column, respectively. Numerical simulations were performed with incoming molecule flux $\phi/k_D = 10^{-5}$ on a regular square lattice.

The deviation from the $N^{1/2}$ scaling observed at $N \sim 10^2$ signals the breakdown, even at optimality, of the low-density regime where the parallel processes of sorting of different species take place approximately independently of each other. For larger N, sorting domains are no longer well separated and molecular mobility is strongly reduced (Fig. 5.3(b)). This can be checked by tracking the diffusive motion of a test molecule which does not interact with any of the molecules of the $i = 1, \ldots, N$ species, except for excluded volume effects. The effective diffusivity of the test molecule, measured empirically from the temporal growth of its mean squared displacement, is observed to decrease when the number N of molecular species increases (Fig. 5.3(b)).

The behavior of the interdomain half distance L, numerically evaluated as a function of g for various values of N, is displayed in Fig. 5.2(e). The shaded area represents the region where the ratio of L to the extraction size $A_E^{1/2}$ is smaller than 1. This region corresponds to a crowding regime where the different species hinder the mobility of each other, thus reducing the sorting efficiency. This is qualitatively confirmed by observing the snapshots of configurations obtained from simulations of the lattice-gas model for different values of N, where the existence of two different regimes, a low-density one and a crowded one, can be clearly distinguished (Fig. 5.4). A convenient measure of the mutual affinity of homotypic molecules, both in a dilute and in a crowded environment, is the effective, macroscopic interaction strength C, which can be easily computed by inverting Eqs. 3.9 and 3.12 [7, 90]. Fig. 5.2(f) shows that C increases monotonically as a function of the microscopic interaction strength

(a)

(b)

(c)

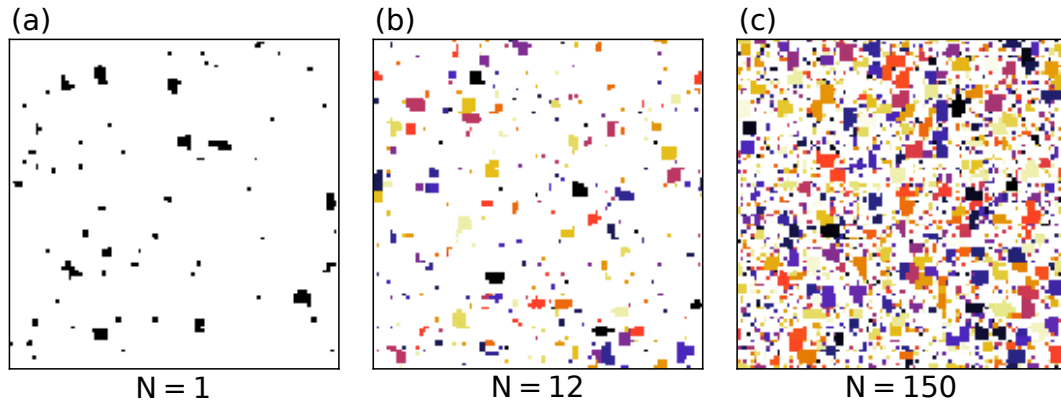N = 1   N = 12   N = 150

Fig. 5.4 Snapshots of the sorting process for increasing values of the number of species N (from left to right). Different molecular species have been marked with different colors. Simulations were performed in the optimal regime at the realistic value $\phi/k_D = 10^{-5}$ of the incoming molecule flux.

g (as previously observed for $N = 1$ in Chapter 4), but decreases monotonically for increasing N, as predicted by the phenomenological theory.

Altogether, these numerical results suggest that up to 10-100 distinct protein cargos might in principle be sorted in parallel, with the highest value attainable only in the optimal regime, while strong crowding effects are expected to impair molecular sorting for larger values of N.

5.2 Sorting species with different mutual affinities

For the symmetric case, where all the molecular species have similar mutual affinities, we found simple scaling laws for the molecular density at the steady state in the optimal region, away from the crowding regime. It is then interesting to investigate to which extent this symmetry requirement is restrictive. To this aim, we here consider the process of sorting of $N = 2$ molecular species, such that their mutual affinities are characterized by independent, and possibly different microscopic interaction strengths g_1 and g_2. By measuring the stationary molecule density ρ, one may look for a global minimum as a function of the two interaction strengths. Fig. 5.5 shows the existence of a single global minimum of the total molecule density ρ for $g_1 = g_2 = g_{opt}$, suggesting that molecular sorting may be most efficient when the distinct sorted molecular species have similar homotypic affinities, and that the
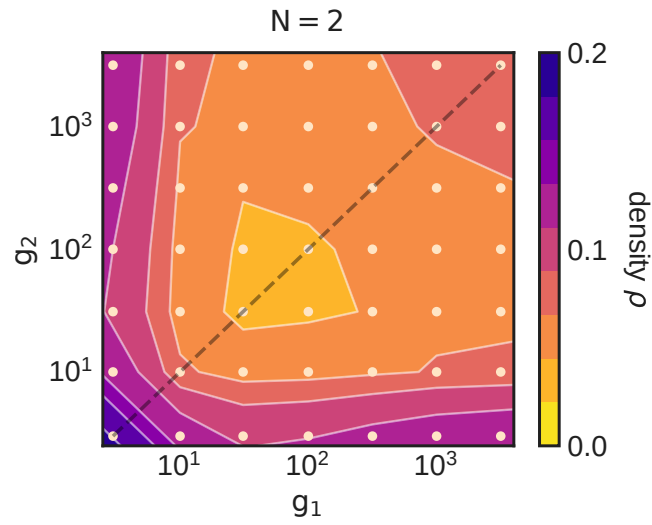
Fig. 5.5 Optimal sorting of two distinct molecular species (lighter region) is obtained for equal homotypic affinities (dashed line $g_1 = g_2$). Simulations were performed at fixed incoming molecule flux $\phi/k_D = 10^{-5}$ on a square lattice. Dots correspond to computed values of ρ, level curves are obtained by linear interpolation of the computed values.

symmetry requirement imposed in the previous treatment may be not too restrictive, as far as the optimal sorting regime is concerned.

Chapter 6

Sorting multivalent molecules

The content of this Chapter is based on the paper under review [91].

An increasing amount of evidence suggests that a crucial factor in a variety of intracellular phase separation processes is valence, which may be defined as the average number of interacting neighbors of a molecule in a phase-separated domain [13, 12, 34]. Experiments have shown that multivalence promotes domain stability [124, 13], and that multivalent protein interactions are responsible for the assembly of endocytic sorting domains [125].

Sec. 6.1 contains a description of how valence can be modeled in this numerical framework, while Sec. 6.2 concerns the effects of valence on the sorting process.

6.1 Modeling valence

A simple way to investigate the role of valence in the present numerical framework is to consider the diffusion of molecules on regular lattices with different coordination numbers z, i.e., on triangular (Fig. 6.1(a)), square (Fig. 6.1(b)), and hexagonal (Fig. 6.1(c)) lattices ($z = 3, 4, 6$). The $z = 8$ case can be implemented by considering the square lattice where nearest neighbors along the diagonals are considered in addition to nearest neighbors along the horizontal and vertical directions. In this lattice-gas framework, the lattice coordination number z can be treated as a proxy of molecular valence. In order to correctly compare the sorting dynamics on lattices
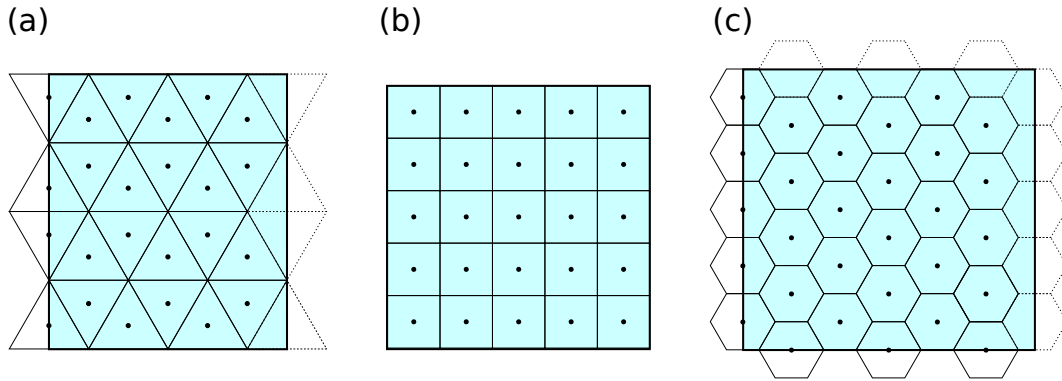
(a)   (b)   (c)

Fig. 6.1 Small periodic regular lattices with $z = 3, 4, 6$, with identical area A_0 of the elementary lattice site and approximately equal total area (light blue). The centers of the sites belonging to each lattice are marked with a black dot.

with different coordination numbers, the microscopic rates were chosen in the following way. The microscopic rate k_D of jump to an empty neighboring site of the lattice can be related to the macroscopic diffusivity D as follows. Consider the diffusion of a single molecule on an otherwise empty lattice, and let $n_x(t)$ be the probability that the molecule occupies site x at time t. At time $t + \delta t$,

$$n_x(t + \delta t) = n_x(t) + k_D \, \delta t \cdot \sum_{y \in \partial x} (n_y - n_x) \tag{6.1}$$

where ∂x is the set of nearest neighbors of x. For the regular lattices with $z = 3, 4, 6$, expanding n_y in a Taylor series centered in x and dividing by δt, Eq. 6.1 tends to the diffusion equation for n_x, with $D = \frac{1}{4} z k_D d^2$, where z is the number of nearest neighbors of x, and d is the distance between the center of neighboring sites. The same procedure applied to the $z = 8$ case with diagonal neighbors gives instead $D = 3 k_D d^2$. For a correct comparison of the sorting dynamics in the presence of different coordination numbers, regular lattices of different coordination number but identical area A_0 of the elementary lattice site (i.e., of identical area per molecule) were used (Fig. 6.1). The total area of each lattice was chosen to contain approximately 100^2 sites, and the microscopic jump rate k_D was rescaled to provide the same value of the macroscopic diffusivity D for all $z = 3, 4, 6, 8$.
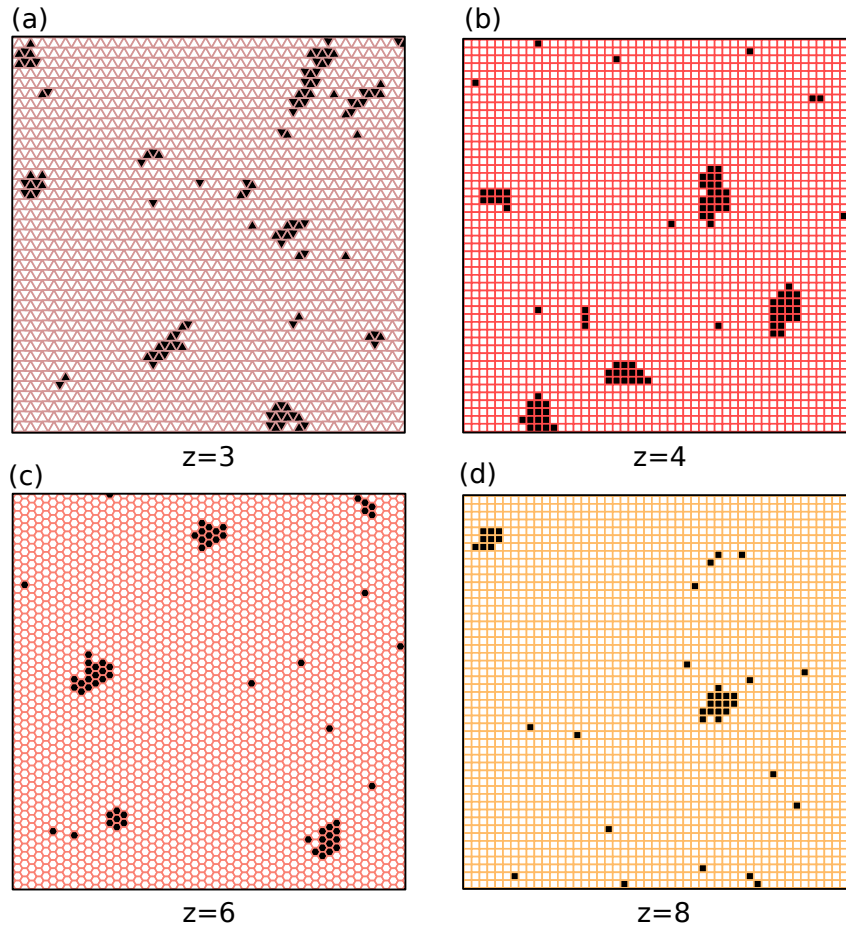
Fig. 6.2 Snapshots of simulated optimal sorting of $N = 1$ molecular species with incoming molecule flux $\phi/k_D = 10^{-5}$ on regular lattices with different coordination number z and equal area A_0 of the elementary lattice site. The panels show enlargements of one quarter of the total system.

6.2   Sorting efficiency

To focus on the dependence of the sorting process on z, the analysis is initially restricted to the case where a single molecular species is sorted ($N = 1$). Snapshots of optimal sorting of a single molecular species on lattices with different coordination number are shown in Fig. 6.2.

Fig. 6.3(a) shows that higher valence implies larger values of the effective, macroscopic interaction strength C at fixed values of the microscopic interaction strength g, consistent with the intuition that molecules of higher valence can more easily aggregate and form phase-separated domains, and that higher valence can
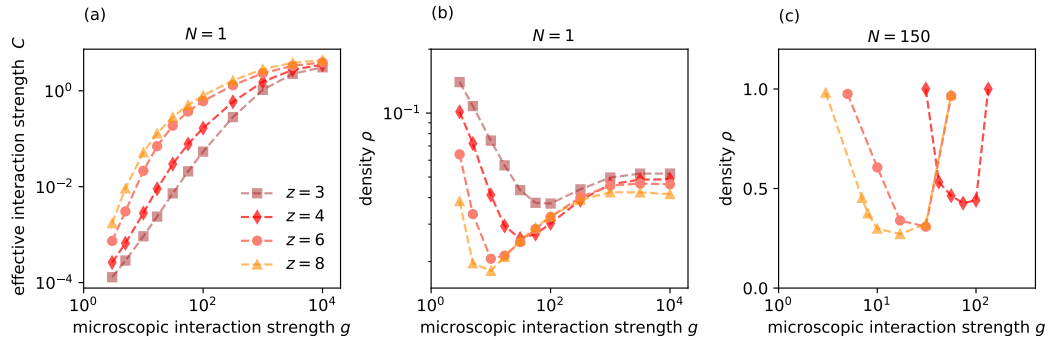
Fig. 6.3 (a) Effective interaction strength C as a function of the microscopic interaction strength g and of the valence z. Higher values of z compensate for smaller g. (b) Total molecule density ρ as a function of the microscopic interaction strength g for different values of the valence z. (c) Density ρ as a function of the interaction strength g for different values of the valence z in a system with $N = 150$ molecular species. Simulations were performed with $\phi/k_D = 10^{-5}$.

compensate for smaller values of the microscopic aggregation strength. This tendency is confirmed by Fig. 6.3(b), showing that optimal sorting (corresponding to the minima of the density curves) is realized at lower values of the microscopic interaction strength g for increasing z. Interestingly, the corresponding optimal values of the stationary molecule density ρ also decrease for increasing z. Perhaps even more importantly, Fig. 6.3(c) shows that the interval of values of g such that the sorting process can take place for high N (here, $N = 150$) significantly widens for increasing z: in this condition, sorting is impossible for $z = 3$ due to molecular crowding, but is instead possible over more then a decade of g values for $z = 6, 8$.

Altogether, these numerical results show that, at least in the present, highly simplified modeling framework, higher valence promotes more efficient sorting. It is interesting to speculate that this may also be true in the case of actual biological systems.

Chapter 7

Coupled lipid-protein dynamics

The formation of phase-separated domains on cell membranes has been investigated in the previous Chapters neglecting the effect of the topological changes induced by vesicle fusion and fission. Here, we propose a simplified model of molecular sorting where these effects are explicitly taken into account. In this model, protein crowding explicitly induces membrane bending and vesicle nucleation. The lipid membrane hosts a binary mixture of proteins which is continuously replenished by the fusion of the membrane with incoming lipid vesicles. The binary mixture phase separates into domains, which in their turn induce the formation of vesicles that are ultimately extracted, thus generating a non-trivial statistical steady state.

In Sec. 7.1, a one-dimensional lattice-gas implementation of the model which explicitly couples membrane and protein dynamics is proposed. Numerical details are provided in Sec. 7.2. Preliminary results from numerical simulations of the model are discussed in Sec. 7.3.

7.1 A coupled lipid-protein model

The lipid membrane is represented as a one-dimensional lattice of N sites with periodic boundary conditions, embedded in a plane, hosting a binary mixture of self-avoiding proteins. A schematic picture of the coupled membrane-protein system is shown in Fig. 7.1(a). Let σ_k be the variable associated to each lattice site, with $k = 1, ..., N$, where $\sigma_k = \pm 1$ if the site is occupied by a protein of one of the two species, $\sigma_k = 0$ otherwise. Protein diffusion is implemented according to the asymmetric

Kawasaki dynamics [126], i.e. a particle can exchange its position with its right nearest neighbour (possibly an empty site), and the tendency of like proteins to aggregate is described by associating to each configuration a Hamiltonian

$$H_{\text{pr}} = -\kappa_A \sum_{k=1}^{N} \sigma_k \sigma_{k+1}, \tag{7.1}$$

where κ_A is the intermolecular interaction strength.

In the time between two protein jumps, the membrane system composed of N nodes $\mathbf{x}_k$ ($k = 1, ..., N$), which is assumed to be immersed in a viscous fluid, evolves according to the overdamped Langevin equation:

$$\dot{\mathbf{x}}_k(t) = -\frac{\partial H_{\text{mem}}(t)}{\partial \mathbf{x}_k} + \boldsymbol{\xi}_k(t). \tag{7.2}$$

To maintain a unitary distance between subsequent nodes $\mathbf{x}_{k-1}$ and $\mathbf{x}_k$, the stochastic force $\boldsymbol{\xi}_k$ describing thermal fluctuations is assumed to have the form $\boldsymbol{\xi}_k = \xi \mathbf{k} \times \mathbf{y}_k$, where $\mathbf{k}$ is the unit outward normal to the plane, $\mathbf{y}_k = \mathbf{x}_k - \mathbf{x}_{k-1}$, $|\mathbf{y}_k| = 1$, ξ is a delta-correlated white Gaussian noise with $\langle \xi(t) \rangle = 0$ and two-point correlation $\langle \xi(t)\xi(t') \rangle = \delta(t - t')\Gamma k_B T$, with Γ a damping coefficient, k_B the Boltzmann constant, and T the temperature. To a given membrane configuration we associate the discrete Helfrich Hamiltonian

$$H_{\text{mem}} = \frac{\kappa_H}{2} \sum_{k=1}^{N} (C_k - C_{0,k})^2, \tag{7.3}$$

where κ_H is the bending rigidity, C_k is the curvature, and the curvature term $C_{0,k}$ takes into account the presence of curvature-inducing proteins on the membrane lattice (Sec. 2.5). The constraint $|\mathbf{y}_k| = 1$ is implemented by the use of the Lagrange multipliers (see App. A for details). The discrete curvature C_k of a configuration is defined as the reciprocal of the radius of the osculating circle shown in Fig. 7.1(b) (see also Sec. 2.5) [127, 128]. By using simple trigonometric relations and the constraint $|\mathbf{y}_k| = 1$, the curvature at node k is given by the expression:

$$C_k = 2 \tan \frac{\theta_k}{2}, \tag{7.4}$$

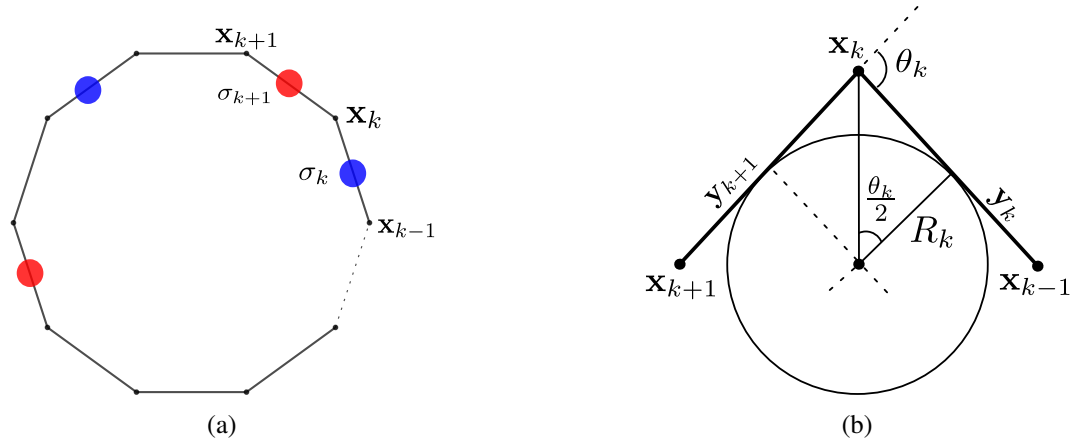where θ_k is the angle shown in Fig. 7.1(b).

Fig. 7.1 (a) Schematic representation of the coupled membrane-protein system. The lipid membrane (black line) is modeled as a 1D lattice with periodic boundary conditions, hosting a binary mixture of proteins (blue and red dots). The dashed line indicates that the number of lattice sites is not fixed. (b) Osculating circle for a given configuration at node k. Each segment, denoted by $\mathbf{y}_k$, has unitary length.

The coupling of lipid membrane and protein dynamics is encoded in the curvature term $C_{0,k}$ of Eq. 7.3, expressed as follows:

$$C_{0,k} = c_{\mathrm{bg}} + c_{\mathrm{pr}}\,\sigma_k \sigma_{k+1} \delta_{\sigma_k \sigma_{k+1}}, \tag{7.5}$$

where c_{bg} is a background curvature, whereas c_{pr} is the curvature induced in protein-crowded regions. As a result of Eq. 7.5, the presence of two like proteins on neighboring sites induces the bending of the lattice.

7.2   Numerical algorithm

Protein diffusion and membrane dynamics

The algorithm is composed of two main parts: protein diffusion and membrane integration. Protein diffusion is implemented by using Gillespie's algorithm [102], consisting of the following steps:

- consider a given protein configuration $\sigma = (\sigma_1, \sigma_2, ..., \sigma_k)$ with $k = 1, ..., N$ at time t;

- make a list of all possible protein jumps with rates $r_k = e^{-\beta \Delta E}$, where $\beta = \log g$, g is the interaction strength, and $\Delta E = (\sigma_{k+2} - \sigma_{k-1})(\sigma_{k+1} - \sigma_k)$ for all k such that $\sigma_k = \pm 1$;

- randomly extract a jump event according to the rates r_k;

- advance the simulation time extracting a random number τ from an exponential distribution with mean $\sum_k r_k$.

The Langevin equation describing the membrane dynamics (Eq. 7.2) is integrated from time t to time $t + \tau$ using the Heun-Stratonovich algorithm, based on the Euler method, with adaptive time step [129, 130].

Vesicle formation and fission

A vesicle is formed and detached from the membrane system as soon as a membrane selfintersection occurs. The algorithm used to detect these events is described in App. B.

Vesicle fusion

Events of vesicle fusion are implemented as follows:

- a lattice node is randomly chosen;

- a vesicle of 10 nodes is attached to that lattice node in a configuration that is tangent to the endosome in that point;

- if the newly attached vesicle does not intersect with any part of the preexisting endosome membrane, the vesicle insertion is accepted, otherwise another lattice node is randomly chosen and the procedure is repeated.

7.3   Stationary state

Numerical simulations of this model were performed starting from an initial membrane configuration of circular shape and size 100 nodes, populated by proteins positioned

on random lattice sites, which was then allowed to relax to the steady state before collecting statistics. The properties of the membrane system (referred to as endosome) and of the extracted vesicles were explored as a function of the background curvature c_{bg} and of the protein-induced curvature c_{pr}. Results obtained at a fixed value of the interaction strength $g = 4$ are displayed in Fig. 7.3. At the stationary state, for low values of c_{bg} and c_{pr}, endosomes have very large sizes, whereas for higher values of both parameters smaller endosome sizes are observed (Fig. 7.3(a)). Fig. 7.3(b) shows how the endosome shape depends on c_{bg} and c_{pr}. This can be quantified in terms of the roundness $4\pi A/p^2$, where p and A denote the perimeter and the area of the endosome, respectively. For $c_{bg} \lesssim 2$ endosomes are observed to be more regular, while for c_{bg} endosomes exhibit star-like shapes, which are unlikely to be assumed by real endosomes. Some representative membrane configurations are reported in Fig. 7.2. In correspondence to the parameter region where giant endosomes are observed, also the size of the extracted vesicle is very large (Fig. 7.3(c)) and the protein residence time is high (Fig. 7.3(d)). Altogether these preliminary results indicate that, in the region where endosomes are characterized by regular shapes, optimal sorting occurs for intermediate values of the protein-induced curvature.
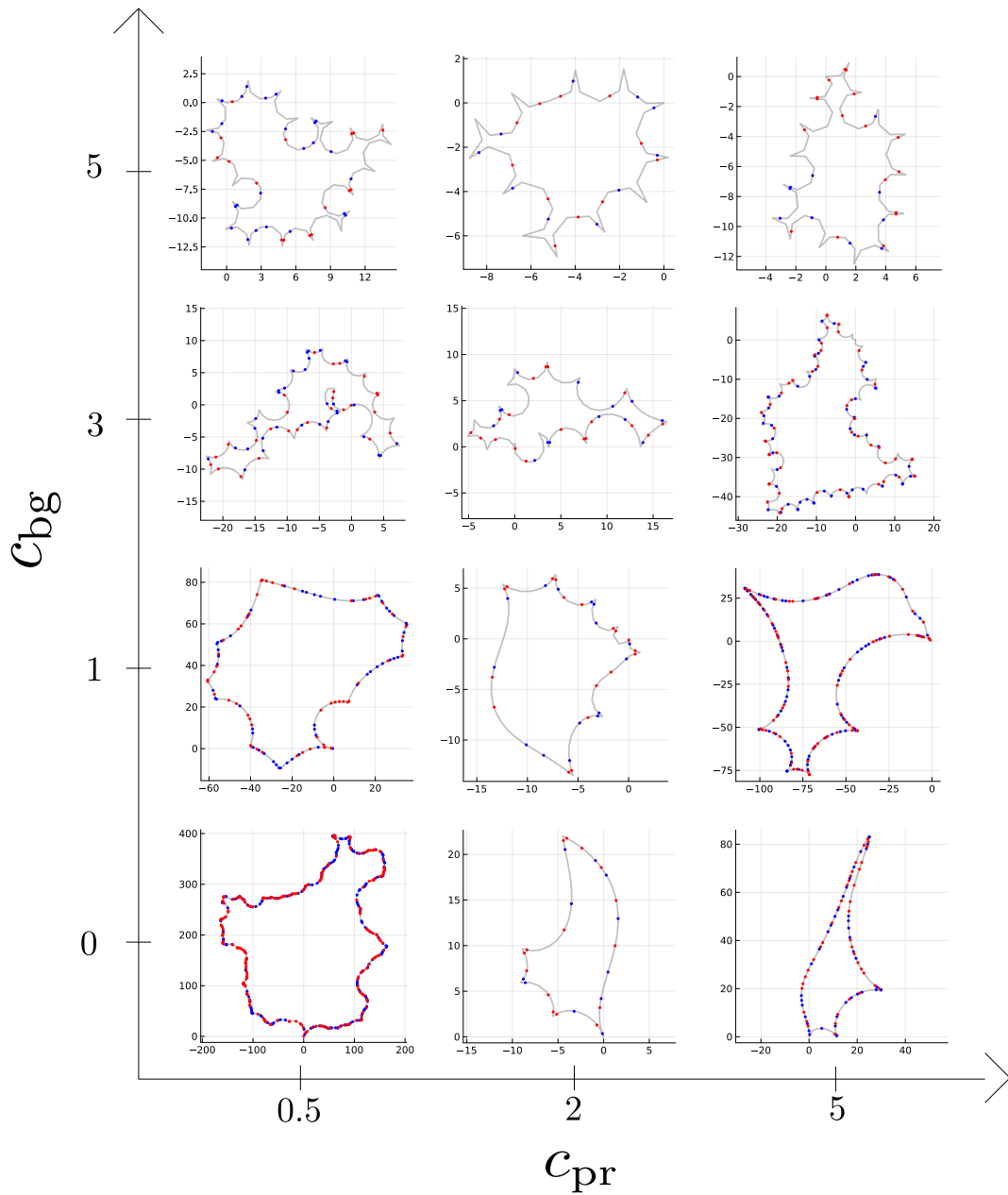
Fig. 7.2 Shape diagram of membrane configurations in the steady state for different values of the background curvature c_{bg} and the curvature induced in protein-crowded regions c_{pr} at fixed interaction strength $g = 4$. Three main regions can be distinguished: (i) for low values of c_{bg} and c_{pr} regular giant endosomes are observed; (ii) for low and intermediate values of c_{bg} endosomes are smaller and characterized by regular shapes; (iii) as c_{bg} is increased, endosomes of irregular shapes are observed.

(a)

(b)

(c)

(d)
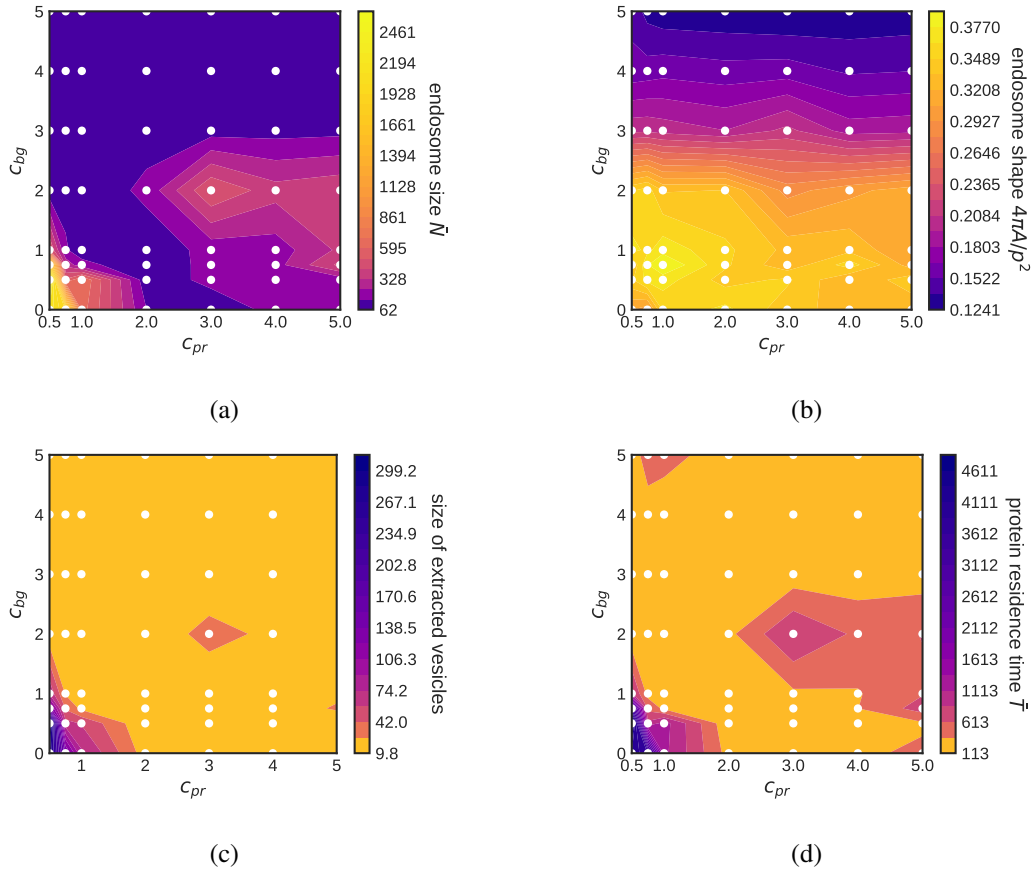
Fig. 7.3 (a) Average endosome size at the steady state. For low values of c_{pr} and c_{bg} the endosome size blows up and, correspondingly, longer and longer times are needed to reach the steady state. (b) The values of c_{bg} controls the endosome roundness. For low c_{bg} endosome are more regular, while for higher c_{bg} endosome have star-like shapes (see also Fig. 7.2. (c) The size of extracted vesicles is close to the size of the inserted vesicles, except for regions where endosomes blow up and giant vesicles are extracted. (d) In the region of round endosomes ($c_{bg} \lesssim 2$) optimal sorting is obtained at intermediate values of c_{pr}. Simulations are performed at fixed interaction strength $g = 4$.

Chapter 8

Conclusions

Achieving spatiotemporal control of the intracellular material is at the core of the proper functioning of the cell, which relies on a multitude of specialized compartments. Such inner structures are generated by a sequence of symmetry-breaking events, whereby each intracellular region is endowed with a specific set of molecular factors that determines its particular physiological function. Evidences have recently shown that phase separation is a main organizing principle behind the partitioning of biomolecules into distinct compartments in the cytoplasm, in the nucleus, and on cell membranes [49, 39, 47, 41, 12, 33, 20]. In addition, the assembly of molecules on cell surfaces has been observed to initiate membrane bending and lead to the formation of lipid vesicles [83–85, 97], each of which encloses specific groups of interacting molecules. Starting from these observations, it has been recently proposed that at the basis of molecular sorting, the crucial process cells exploit to maintain the specific biochemical identity of each compartment, is the coupling of two fundamental mechanisms: molecule phase separation into localized domains and vesicle nucleation induced by domain formation.

A central notion of the theory of phase separation is that only domains larger than a critical size A_c are able to grow indefinitely, while smaller domains tend to be dissolved. In combination with a contextual process of domain extraction at a larger scale $A_E > A_c$, this introduces a sort of "physical checkpoint", such that only domains that are able to reach the "critical mass" A_c can drive extraction events, and are thus "productive". This scenario is consistent with experimental observations where, in addition to "productive" long-lived domains that grow into vesicles that

are ultimately extracted from the membrane, a large number of short-lived, small domains, which tend to disassemble and ultimately disappear, is also detected. The existence of such a "physical checkpoint" is reflected in the particular shape of the size distribution for productive domains (Eq. 3.5), which exhibits a maximum at sizes of the order of the critical size A_c, a slowly (logarithmically) decaying intermediate region, followed by a non-universal decaying tail at scales larger than the extraction threshold A_E (Fig. 4.6(a), blue histogram). On the other hand, the existence of a biochemical checkpoint has also been postulated in this regard [106, 107]. It would be quite interesting to further investigate the relation between these two effects.

It is worth observing here that in the actual biophysical process, a wealth of different biomolecular species takes place in the formation and stabilization of sorting domains. In the theoretical model, the complex interplay between these different species is effectively encoded into the value of the single dimensionless interaction parameter g. Intriguingly, even such a highly simplified abstract model, founded on basic notions from the theory of phase separation, is able to capture relevant features of the real process. This yields support to the hypothesis that endocytic sorting is driven by an underlying phase separation process.

Although, for simplicity, a single molecular species was initially considered in this framework, the sorting process taking place on lipid membranes actually consists in the demixing of a myriad of different molecules. Therefore, it is interesting to extend the model in order to consider a heterogeneous pool of molecules. Analytical arguments provide a qualitative indication of the number of species which can be sorted in parallel, suggesting that sorting of a large number of species is possible. A more precise estimate is obtained by numerical investigations, showing that at fixed incoming molecular flux the average time molecules spend on the membrane increases with the number of species, following the simple scaling law $N^{1/2}$. Due to molecular crowding effects, this scaling law is observed to break down at $N \approx 10^2$. In this crowding regime, molecular mobility is reduced and domains are no longer well-separated.

Motivated by recent experimental evidences pointing out the essential role of multivalent proteins in driving endocytic sorting and favoring domain stability [13, 124, 125], the impact of molecular valence on the sorting process has been also studied in this highly simplified framework. The implementation of regular lattices with different coordination numbers allowed to simulate the aggregation of molecules

with different valence, showing that the sorting process results to be most efficient for higher-valence molecules.

The last part of this thesis work is dedicated to a preliminary study of the interplay between molecules and lipid membrane. It is indeed reported in several works that protein crowding drives membrane bending [85, 83, 84] and, conversely, membrane dynamics affects the formation of protein assemblies. Changes in membrane curvature can also lead to the formation of vesicles that, once mature, detach from the donor membrane and can attach to a target membrane releasing their content. To model the coupling of protein and membrane dynamics and to simulate the topological changes the lipid membrane undergoes, a new model is proposed and preliminarily investigated through numerical simulations on a one-dimensional fluctuating topologically-varying lattice.

Finally, it should be kept in mind that other computational approaches could be used to investigate the process of molecular sorting. It would be interesting, for instance, to combine the scenario emerging from our minimal model with investigations performed through Molecular Dynamics simulations, attaining a more comprehensive description of molecular sorting.

References

[1] Alexander Johnson, Julian Lewis, and Bruce ALBERTS. Molecular biology of the cell. Garland Science New York, NY, USA:, 2002.

[2] Steven Boeynaems, Simon Alberti, Nicolas L. Fawzi, Tanja Mittag, Magdalini Polymenidou, Frederic Rousseau, Joost Schymkowitz, James Shorter, Benjamin Wolozin, Ludo Van Den Bosch, Peter Tompa, and Monika Fuxreiter. Protein phase separation: A new phase in cell biology. Trends in Cell Biology, 28(6):420–435, June 2018.

[3] J. Halatek, F. Brauns, and E. Frey. Self-organization principles of intracellular pattern formation. Philosophical Transactions of the Royal Society B: Biological Sciences, 373(1747):20170107, apr 2018.

[4] E. Floris, A. Piras, L. Dall'Asta, A. Gamba, E. Hirsch, and C.C. Campa. Physics of compartmentalization: How phase separation and signaling shape membrane and organelle identity. Comput. Struct. Biotechnol. J., 19:3225–3233, 2021.

[5] Ira Mellman and W James Nelson. Coordinated protein sorting, targeting and distribution in polarized cells. Nature reviews Molecular cell biology, 9(11):833–845, 2008.

[6] Peter J Cullen and Florian Steinberg. To degrade or not to degrade: mechanisms and significance of endocytic recycling. Nature reviews Molecular cell biology, 19(11):679–696, 2018.

[7] M. Zamparo, D. Valdembri, G. Serini, I.V. Kolokolov, V.V. Lebedev, L. Dall'Asta, and A. Gamba. Optimality in self-organized molecular sorting. Phys. Rev. Lett., 126:088101–1–7, 2021.

[8] Sofia Rodriguez-Gallardo, Kazuo Kurokawa, Susana Sabido-Bozo, Alejandro Cortes-Gomez, Atsuko Ikeda, Valeria Zoni, Auxiliadora Aguilera-Romero, Ana Maria Perez-Linero, Sergio Lopez, Miho Waga, et al. Ceramide chain length–dependent protein sorting into selective endoplasmic reticulum exit sites. Science advances, 6:eaba8237, 2020.

[9] Sofia Rodriguez-Gallardo, Kazuo Kurokawa, Susana Sabido-Bozo, Alejandro Cortes-Gomez, Ana Maria Perez-Linero, Auxiliadora Aguilera-Romero,

Sergio Lopez, Miho Waga, Akihiko Nakano, and Manuel Muñiz. Assay for dual cargo sorting into endoplasmic reticulum exit sites imaged by 3D Super-resolution Confocal Live Imaging Microscopy (SCLIM). Plos one, 16(10):e0258111, 2021.

[10] Yutaro Shimizu, Junpei Takagi, Emi Ito, Yoko Ito, Kazuo Ebine, Yamato Komatsu, Yumi Goto, Mayuko Sato, Kiminori Toyooka, Takashi Ueda, et al. Cargo sorting zones in the trans-golgi network visualized by super-resolution confocal live imaging microscopy in plants. Nature communications, 12:1–14, 2021.

[11] Natalia Gomez-Navarro and Elizabeth Miller. Protein sorting at the er–golgi interface. Journal of Cell Biology, 215:769–778, 2016.

[12] Sudeep Banjade and Michael K Rosen. Phase transitions of multivalent proteins can promote clustering of membrane receptors. elife, 3:e04123, 2014.

[13] Pilong Li, Sudeep Banjade, Hui-Chun Cheng, Soyeon Kim, Baoyu Chen, Liang Guo, Marc Llaguno, Javoris V Hollingsworth, David S King, Salman F Banani, et al. Phase transitions in the assembly of multivalent signalling proteins. Nature, 483(7389):336–340, 2012.

[14] Sudeep Banjade, Lu Zhu, Jeffrey R Jorgensen, Sho W Suzuki, and Scott D Emr. Recruitment and organization of escrt-0 and ubiquitinated cargo via condensation. Sci. Adv., 8:eabm5149, 2022.

[15] Rong Li and Bruce Bowerman. Symmetry breaking in biology. Cold Spring Harb Perspect Biol, 2(3):a003475, March 2010.

[16] Zsolt G Venkei and Yukiko M Yamashita. Emerging mechanisms of asymmetric stem cell division. Journal of Cell Biology, 217(11):3785–3795, 2018.

[17] Anne J Ridley, Martin A Schwartz, Keith Burridge, Richard A Firtel, Mark H Ginsberg, Gary Borisy, J Thomas Parsons, and Alan Rick Horwitz. Cell migration: integrating signals from front to back. Science, 302(5651):1704–1709, 2003.

[18] Annette Shewan, Dennis J Eastburn, and Keith Mostov. Phosphoinositides in cell architecture. Cold Spring Harbor perspectives in biology, 3(8):a004796, 2011.

[19] A. Gamba, A. de Candia, S. Di Talia, A. Coniglio, F. Bussolino, and G. Serini. Diffusion limited phase separation in eukaryotic chemotaxis. Proc. Natl. Acad. Sci. U.S.A., 102:16927–32, 2005.

[20] A Gamba, I Kolokolov, V Lebedev, and G Ortenzi. Patch coalescence as a mechanism for eukaryotic directional sensing. Physical review letters, 99(15):158101, 2007.

[21] Andrea Veglio, A Gamba, Mario Nicodemi, Federico Bussolino, and Guido Serini. Symmetry breaking mechanism for epithelial cell polarization. Physical Review E, 80(3):031919, 2009.

[22] Matteo Semplice, Andrea Veglio, Giovanni Naldi, Guido Serini, and Andrea Gamba. A bistable model of cell polarity. PLoS One, 7:e30977, 2012.

[23] Michel Bornens. Organelle positioning and cell polarity. Nature Reviews Molecular Cell Biology, 9(11):874–886, 2008.

[24] Smita Yadav, Sapna Puri, and Adam D Linstedt. A primary role for Golgi positioning in directed secretion, cell polarity, and wound healing. Molecular biology of the cell, 20(6):1728–1736, 2009.

[25] Enrique Rodriguez-Boulan and Ian G Macara. Organization and execution of the epithelial polarity programme. Nature reviews Molecular cell biology, 15(4):225–242, 2014.

[26] Yamini Ravichandran, Bruno Goud, and Jean-Baptiste Manneville. The Golgi apparatus and cell polarity: Roles of the cytoskeleton, the Golgi matrix, and Golgi membranes. Current opinion in cell biology, 62:104–113, 2020.

[27] Yusong Guo, Daniel W Sirkis, and Randy Schekman. Protein sorting at the trans-Golgi network. Annual review of cell and developmental biology, 30:169–206, 2014.

[28] Christine Kienzle and Julia von Blume. Secretory cargo sorting at the trans-golgi network. Trends in cell biology, 24(10):584–593, 2014.

[29] Benedetta Cerruti, Alberto Puliafito, Annette M. Shewan, Wei Yu, Alexander N. Combes, Melissa H. Little, Federica Chianale, Luca Primo, Guido Serini, Keith E. Mostov, Antonio Celani, and Andrea Gamba. Polarity, cell division, and out-of-equilibrium dynamics control the growth of epithelial structures. J Cell Biol, 203:359–72, 2013.

[30] Xiaona Fang, Karsten Kruse, Ting Lu, and Jin Wang. Nonequilibrium physics in biology. Reviews of Modern Physics, 91(4):045004, 2019.

[31] Marina Feric, Nilesh Vaidya, Tyler S Harmon, Diana M Mitrea, Lian Zhu, Tiffany M Richardson, Richard W Kriwacki, Rohit V Pappu, and Clifford P Brangwynne. Coexisting liquid phases underlie nucleolar subcompartments. Cell, 165(7):1686–1697, 2016.

[32] Avinash Patel et al. A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. Cell, 162(5):1066–1077, 2015.

[33] Xiaolei Su, Jonathon A. Ditlev, Enfu Hui, Wenmin Xing, Sudeep Banjade, Julia Okrut, David S. King, Jack Taunton, Michael K. Rosen, and Ronald D. Vale. Phase separation of signaling molecules promotes T cell receptor signal transduction. Science, 352(6285):595–599, Apr 2016.

[34] Yongdae Shin and Clifford P Brangwynne. Liquid phase condensation in cell physiology and disease. <u>Science</u>, 357(6357):eaaf4382, 2017.

[35] Andrew S Lyon, William B Peeples, and Michael K Rosen. A framework for understanding the functions of biomolecular condensates across scales. <u>Nat Rev Mol Cell Biol</u>, November 2021.

[36] Titus M Franzmann, Marcus Jahnel, Andrei Pozniakovsky, Julia Mahamid, Alex S Holehouse, Elisabeth Nüske, Doris Richter, Wolfgang Baumeister, Stephan W Grill, Rohit V Pappu, et al. Phase separation of a yeast prion protein promotes cellular fitness. <u>Science</u>, 359(6371), 2018.

[37] Salman F. Banani, Hyun O. Lee, Anthony A. Hyman, and Michael K. Rosen. Biomolecular condensates: organizers of cellular biochemistry. <u>Nature Reviews Molecular Cell Biology</u>, 18(5):285–298, February 2017.

[38] Satomi Matsuoka, Tatsuo Shibata, and Masahiro Ueda. Asymmetric PTEN distribution regulated by spatial heterogeneity in membrane-binding state transitions. <u>PLoS Comput Biol</u>, 9(1):e1002862, 2013.

[39] Joshua R Wheeler, Tyler Matheny, Saumya Jain, Robert Abrisch, and Roy Parker. Distinct stages in stress granule assembly and disassembly. <u>eLife</u>, 5, 09 2016.

[40] Menglong Zeng, Yuan Shang, Yoichi Araki, Tingfeng Guo, Richard L Huganir, and Mingjie Zhang. Phase transition in postsynaptic densities underlies formation of synaptic complexes and synaptic plasticity. <u>Cell</u>, 166(5):1163–1175, 2016.

[41] Joel Berry, Stephanie C. Weber, Nilesh Vaidya, Mikko Haataja, and Clifford P. Brangwynne. RNA transcription modulates phase transition-driven nuclear body assembly. <u>Proceedings of the National Academy of Sciences</u>, 112(38):E5237–E5245, September 2015.

[42] Audrey S. Howell, Meng Jin, Chi-Fang Wu, Trevin R. Zyla, Timothy C. Elston, and Daniel J. Lew. Negative feedback enhances robustness in the yeast polarity establishment circuit. <u>Cell</u>, 149:322–33, 2012.

[43] Joshua A Riback, Christopher D Katanski, Jamie L Kear-Scott, Evgeny V Pilipenko, Alexandra E Rojek, Tobin R Sosnick, and D Allan Drummond. Stress-triggered phase separation is an adaptive, evolutionarily tuned response. <u>Cell</u>, 168(6):1028–1040, 2017.

[44] Tetsuro Hirose, Giorgio Virnicchi, Akie Tanigawa, Takao Naganuma, Ruohan Li, Hiroshi Kimura, Takahide Yokoi, Shinichi Nakagawa, Marianne Bénard, Archa H Fox, et al. NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies. <u>Molecular biology of the cell</u>, 25(1):169–183, 2014.

[45] Aaron Kaplan and Leonora Reinhold. CO2 concentrating mechanisms in photosynthetic microorganisms. Annual review of plant biology, 50(1):539–570, 1999.

[46] Luke M Oltrogge, Thawatchai Chaijarasphong, Allen W Chen, Eric R Bolin, Susan Marqusee, and David F Savage. Multivalent interactions between CsoS2 and Rubisco mediate α-carboxysome formation. Nature structural & molecular biology, 27(3):281–287, 2020.

[47] David Zwicker, Markus Decker, Steffen Jaensch, Anthony A Hyman, and Frank Jülicher. Centrosomes are autocatalytic droplets of pericentriolar material organized by centrioles. Proceedings of the National Academy of Sciences, 111(26):E2636–E2645, 2014.

[48] Jeffrey B Woodruff, Beatriz Ferreira Gomes, Per O Widlund, Julia Mahamid, Alf Honigmann, and Anthony A Hyman. The centrosome is a selective condensate that nucleates microtubules by concentrating tubulin. Cell, 169(6):1066–1077, 2017.

[49] C. P. Brangwynne, C. R. Eckmann, D. S. Courson, A. Rybarska, C. Hoege, J. Gharakhani, F. Julicher, and A. A. Hyman. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. Science, 324(5935):1729–32, 2009.

[50] Carsten Hoege and Anthony A Hyman. Principles of PAR polarity in caenorhabditis elegans embryos. Nature reviews Molecular cell biology, 14(5):315–322, 2013.

[51] A. Gamba, I. Kolokolov, V. Lebedev, and G. Ortenzi. Universal features of cell polarization processes. J Stat Mech, page P02019, 2009.

[52] Elisa Floris, Andrea Piras, Luca Dall'Asta, Andrea Gamba, Emilio Hirsch, and Carlo C. Campa. Physics of compartmentalization: How phase separation and signaling shape membrane and organelle identity. Comp. Struct. Biotechn. J., 19:3225–3233, 2021.

[53] Alice Cezanne, Janelle Lauer, Anastasia Solomatina, Ivo F Sbalzarini, and Marino Zerial. A non-linear system patterns Rab5 GTPase on the membrane. Elife, 9:e54434, 2020.

[54] Kurt Binder and Dietrich Stauffer. Statistical theory of nucleation, condensation and coagulation. Advances in Physics, 25(4):343–396, 1976.

[55] Alan J Bray. Theory of phase-ordering kinetics. Advafnces in Physics, 51(2):481–587, 2002.

[56] Vinod Wadhawan and Sanjay Puri, editors. Kinetics of phase transitions. CRC Press, 2009.

[57] Jacob N. Israelachvili. Intermolecular and Surface Forces, Third Edition: Revised Third Edition. Academic Press, 3 edition, 2011.

[58] D Leckband and J Israelachvili. Intermolecular forces in biology. Q Rev Biophys, 34(2):105–267, May 2001.

[59] D Chandler, J D Weeks, and H C Andersen. Van der Waals picture of liquids, solids, and phase transformations. Science, 220(4599):787–94, May 1983.

[60] E.M. Lifshitz and L.P. Pitaevskii. Physical Kinetics, volume 10 of Course of Theoretical Physics. Pergamon Press, first edition, 1981.

[61] L.D. Landau and E.M. Lifshitz. Statistical Physics (Part I), volume 5 of Course of Theoretical Physics. Pergamon Press, third edition, 1980.

[62] Vitaly V Slezov. Kinetics of first order phase transitions. John Wiley & Sons, 2009.

[63] Pavel L Krapivsky, Sidney Redner, and Eli Ben-Naim. A kinetic view of statistical physics. Cambridge University Press, 2010.

[64] IM Lifshitz and VV Slezov. Kinetics of diffusive decomposition of supersaturated solid solutions. Soviet Physics JETP, 35(8):331–339, 1959.

[65] Davide Marenduzzo. An introduction to the statistical physics of active matter: motility-induced phase separation and the "generic instability" of active gels. The European Physical Journal Special Topics, 225(11):2065–2077, 2016.

[66] Katja Lindenberg, Ralf Metzler, and Gleb Oshanin. Chemical Kinetics: Beyond the Textbook. World Scientific Publishing Europe Ltd, 2019.

[67] David Zwicker, Anthony A Hyman, and Frank Juelicher. Suppression of Ostwald ripening in active emulsions. Physical Review E, 92(1):012317, 2015.

[68] Yoshiyuki Arai, Tatsuo Shibata, Satomi Matsuoka, Masayuki J Sato, Toshio Yanagida, and Masahiro Ueda. Self-organization of the phosphatidylinositol lipids signaling system for random cell migration. Proceedings of the National Academy of Sciences, 107(27):12399–12404, 2010.

[69] Nathan W Goehring, Philipp Khuc Trong, Justin S Bois, Debanjan Chowdhury, Ernesto M Nicola, Anthony A Hyman, and Stephan W Grill. Polarization of PAR proteins by advective triggering of a pattern-forming system. Science, 334(6059):1137–1141, 2011.

[70] Jayme M Johnson, Meng Jin, and Daniel J Lew. Symmetry breaking and the establishment of cell polarity in budding yeast. Current opinion in genetics & development, 21(6):740–746, 2011.

[71] Chi-Fang Wu, Jian-Geng Chiou, Maria Minakova, Benjamin Woods, Denis Tsygankov, Trevin R Zyla, Natasha S Savage, Timothy C Elston, and Daniel J Lew. Role of competition between polarity sites in establishing a unique front. Elife, 4:e11611, 2015.

[72] Marco Zamparo, F Chianale, Claudio Tebaldi, M Cosentino-Lagomarsino, M Nicodemi, and A Gamba. Dynamic membrane patterning, signal localization and polarity in living cells. Soft Matter, 11(5):838–849, 2015.

[73] M. Otsuji, S. Ishihara, C. Co, K. Kaibuchi, A. Mochizuki, and S. Kuroda. A mass conserved reaction-diffusion system captures properties of cell polarity. PLoS Comput Biol, 3(6):e108, 2007.

[74] C. Beta, G. Amselem, and E. Bodenschatz. A bistable mechanism for directional sensing. New J Phys, 10:083015, 2008.

[75] Y. Mori, A. Jilkine, and L.f Edelstein-Keshet. Wave-pinning and cell polarity from a bistable reaction-diffusion system. Biophys J, 94:3684, 2008.

[76] Vitaly I Kalikmanov. Classical nucleation theory. In Nucleation theory, pages 17–41. Springer, 2012.

[77] Richard Becker and Werner Döring. Kinetische behandlung der keimbildung in übersättigten dämpfen. Ann. Phys., 416(8):719–752, 1935.

[78] Ya. B. Zeldovich. On the Theory of New Phase Formation: Cavitation. Acta Physicochem. USSR, 18:1, 1943.

[79] Jeanne C Stachowiak, Frances M Brodsky, and Elizabeth A Miller. A cost–benefit analysis of the physical mechanisms of membrane curvature. Nature cell biology, 15(9):1019–1027, 2013.

[80] Iris K Jarsch, Frederic Daste, and Jennifer L Gallop. Membrane curvature in cell biology: An integration of molecular mechanisms. Journal of Cell Biology, 214(4):375–387, 2016.

[81] Harvey T McMahon and Jennifer L Gallop. Membrane curvature and mechanisms of dynamic cell membrane remodelling. Nature, 438(7068):590–596, 2005.

[82] Jeanne C Stachowiak, Carl C Hayden, and Darryl Y Sasaki. Steric confinement of proteins on lipid membranes can drive curvature and tubulation. Proceedings of the National Academy of Sciences, 107(17):7781–7786, 2010.

[83] Jeanne C Stachowiak, Eva M Schmid, Christopher J Ryan, Hyoung Sook Ann, Darryl Y Sasaki, Michael B Sherman, Phillip L Geissler, Daniel A Fletcher, and Carl C Hayden. Membrane bending by protein–protein crowding. Nature cell biology, 14(9):944–949, 2012.

[84] Wilton T Snead, Carl C Hayden, Avinash K Gadok, Chi Zhao, Eileen M Lafer, Padmini Rangamani, and Jeanne C Stachowiak. Membrane fission by protein crowding. Proceedings of the National Academy of Sciences, 114(16):E3258–E3267, 2017.

[85] Feng Yuan, Haleh Alimohamadi, Brandon Bakka, Andrea N Trementozzi, Kasey J Day, Nicolas L Fawzi, Padmini Rangamani, and Jeanne C Stachowiak. Membrane bending by protein phase separation. Proceedings of the National Academy of Sciences, 118(11):e2017435118, 2021.

[86] Udo Seifert. Configurations of fluid membranes and vesicles. Advances in physics, 46(1):13–137, 1997.

[87] Peter B Canham. The minimum energy of bending as a possible explanation of the biconcave shape of the human red blood cell. Journal of theoretical biology, 26(1):61–81, 1970.

[88] Wolfgang Helfrich. Elastic properties of lipid bilayers: theory and possible experiments. Zeitschrift für Naturforschung c, 28(11-12):693–703, 1973.

[89] N Ramakrishnan, PB Sunil Kumar, and Ravi Radhakrishnan. Mesoscale computational studies of membrane bilayer remodeling by curvature-inducing proteins. Physics reports, 543(1):1–60, 2014.

[90] Elisa Floris, Andrea Piras, Francesco Saverio Pezzicoli, Marco Zamparo, Luca Dall'Asta, and Andrea Gamba. Phase separation and critical size in molecular sorting. Phys. Rev. E, 106:044412, Oct 2022.

[91] Andrea Piras, Elisa Floris, Luca Dall'Asta, and Andrea Gamba. Sorting of multiple molecular species on cell membranes. arXiv preprint arXiv:2301.05966, 2023.

[92] Owen Pornillos, Jennifer E Garrus, and Wesley I Sundquist. Mechanisms of enveloped RNA virus budding. Trends Cell Biol., 12:569–579, 2002.

[93] Jeremy S Rossman and Robert A Lamb. Influenza virus assembly and budding. Virology, 411:229–236, 2011.

[94] Prabuddha Sengupta and Jennifer Lippincott-Schwartz. Revisiting membrane microdomains and phase separation: a viral perspective. Viruses, 12(7):745, 2020.

[95] Balindile B Motsa and Robert V Stahelin. Lipid–protein interactions in virus assembly and budding from the host cell plasma membrane. Bioch. Soc. Trans., 49:1633–1641, 2021.

[96] Joel Berry, Clifford P Brangwynne, and Mikko Haataja. Physical principles of intracellular organization via active and passive phase transitions. Rep. Prog. Phys., 81:046601, 2018.

[97] David J Busch, Justin R Houser, Carl C Hayden, Michael B Sherman, Eileen M Lafer, and Jeanne C Stachowiak. Intrinsically disordered proteins drive membrane curvature. Nature communications, 6(1):7875, 2015.

[98] N. S. Gov. Guided by curvature: shaping cells by coupling curved membrane proteins and cytoskeletal forces. Philosophical Transactions of the Royal Society B: Biological Sciences, 373(1747):20170115, 2018.

[99] Sara Sigismund, Stefano Confalonieri, Andrea Ciliberto, Simona Polo, Giorgio Scita, and Pier Paolo Di Fiore. Endocytosis and signaling: cell logistics shape the eukaryotic cell plan. Physiol. Rev., 92:273–366, 2012.

[100] Marco Zamparo, Luca Dall'Asta, and Andrea Gamba. On the mean residence time in stochastic lattice-gas models. Journal of Statistical Physics, 174(1):120–134, 2019.

[101] Giovanni B. Brandani, Marieke Schor, Cait E. MacPhee, Helmut Grubmüller, Ulrich Zachariae, and Davide Marenduzzo. Quantifying disorder through conditional entropy: An application to fluid mixing. Plos One, 8:e65617, 2013.

[102] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of computational physics, 22(4):403–434, 1976.

[103] F den Hollander, Enzo Olivieri, and Elisabetta Scoppola. Metastability and nucleation for conservative dynamics. J. Math Phys., 41(3):1424–1498, 2000.

[104] Ladislaus Farkas. Keimbildungsgeschwindigkeit in übersättigten dämpfen. Z. Phys. Chem., 125(1):236–242, 1927.

[105] Jürn Schmelzer, Gerd Röpke, and Frank-Peter Ludwig. Nuclear multifragmentation processes and nucleation theory. Phys. Rev. C, 55(4):1917, 1997.

[106] Dinah Loerke, Marcel Mettlen, Defne Yarar, Khuloud Jaqaman, Henry Jaqaman, Gaudenz Danuser, and Sandra L Schmid. Cargo and dynamin regulate clathrin-coated pit maturation. PLoS Biol., 7(3):e1000057, mar 2009.

[107] François Aguet, Costin N. Antonescu, Marcel Mettlen, Sandra L. Schmid, and Gaudenz Danuser. Advances in analysis of low signal-to-noise images link dynamin and AP2 to the functions of an endocytic checkpoint. Dev. Cell, 26(3):279–291, aug 2013.

[108] Zuzana Kadlecova, Stephanie J. Spielman, Dinah Loerke, Aparna Mohanakrishnan, Dana Kim Reed, and Sandra L. Schmid. Regulation of clathrin-mediated endocytosis by hierarchical allosteric activation of AP2. J. Cell Biol., 216(1):167–179, 12 2016.

[109] Alexandre Grassart, Aaron T. Cheng, Sun Hae Hong, Fan Zhang, Nathan Zenzer, Yongmei Feng, David M. Briner, Gregory D. Davis, Dmitry Malkov, and David G. Drubin. Actin and dynamin2 dynamics and interplay during clathrin-mediated endocytosis. J. Cell Biol., 205:721–735, 2014.

[110] Kangmin He, Eli Song, Srigokul Upadhyayula, Song Dang, Raphael Gaudin, Wesley Skillern, Kevin Bu, Benjamin R. Capraro, Iris Rapoport, Ilja Kusters, Minghe Ma, and Tom Kirchhausen. Dynamics of Auxilin 1 and GAK in clathrin-mediated traffic. J. Cell Biol., 219:e201908142, 01 2020.

[111] Xinxin Wang, Zhiming Chen, Marcel Mettlen, Jungsik Noh, Sandra L Schmid, and Gaudenz Danuser. DASC, a sensitive classifier for measuring discrete early stages in clathrin-mediated endocytosis. eLife, 9:e53686, apr 2020.

[112] Martin Lehmann, Ilya Lukonin, Frank Noé, Jan Schmoranzer, Cecilia Clementi, Dinah Loerke, and Volker Haucke. Nanoscale coupling of endocytic pit growth and stability. Sci. Adv., 5(11):eaax5775, 2019.

[113] Marcelo Ehrlich, Werner Boll, Antoine van Oijen, Ramesh Hariharan, Kartik Chandran, Max L. Nibert, and Tomas Kirchhausen. Endocytosis by random initiation and stabilization of clathrin-coated pits. Cell, 118:591–605, sep 2004.

[114] A-L Barabási, Harry Eugene Stanley, et al. Fractal concepts in surface growth. Cambridge University Press, 1995.

[115] Marcel Mettlen, Dinah Loerke, Defne Yarar, Gaudenz Danuser, and Sandra L. Schmid. Cargo- and adaptor-specific mechanisms regulate clathrin-mediated endocytosis. J. Cell Biol., 188:919–933, 2010.

[116] Anthony A. Hyman, Christoph A. Weber, and Frank Jülicher. Liquid-Liquid Phase Separation in Biology. Ann. Rev. Cell Dev. Biol., 30:39–58, 2014.

[117] Atsuyuki Okabe, Barry Boots, Kokichi Sugihara, Sung Nok Chiu, and D. G. Kendall. Spatial Tessellations. Wiley Series in Probability and Statistics, 2000.

[118] Seunghwa Ryu and Wei Cai. Numerical tests of nucleation theories for the ising models. Phys. Rev. E, 82(1):011603, 2010.

[119] Allen P. Liu, François Aguet, Gaudenz Danuser, and Sandra L. Schmid. Local clustering of transferrin receptors promotes clathrin-coated pit initiation. J. Cell Biol., 191:1381–1393, 2010.

[120] Shawn M Ferguson and Pietro De Camilli. Dynamin, a membrane-remodelling gtpase. Nat. Rev. Mol. Cell Biol., 13(2):75–88, 2012.

[121] Sun Hae Hong, Christa L. Cortesio, and David G. Drubin. Machine-Learning-Based Analysis in Genome-Edited Cells Reveals the Efficiency of Clathrin-Mediated Endocytosis. Cell Rep., 12(12):2121–2130, Sep 2015.

[122] Nicolas Destainville. An alternative scenario for the formation of specialized protein nano-domains (cluster phases) in biomembranes. EPL, 91:58001, 2010.

[123] Nicolas Meilhac and Nicolas Destainville. Clusters of proteins in biomembranes: insights into the roles of interaction potential shapes and of protein diversity. J. Phys. Chem. B, 115:7190–7199, 2011.

[124] Salman F Banani, Allyson M Rice, William B Peeples, Yuan Lin, Saumya Jain, Roy Parker, and Michael K Rosen. Compositional control of phase-separated cellular bodies. Cell, 166(3):651–663, 2016.

[125] Kasey J. Day, Grace Kago, Liping Wang, J. Blair Richter, Carl C. Hayden, Eileen M. Lafer, and Jeanne C. Stachowiak. Liquid-like protein interactions catalyse assembly of endocytic vesicles. Nat. Cell Biol., 23:366–376, Apr 2021.

[126] Kyozi Kawasaki. Diffusion constants near the critical point for time-dependent ising models. i. Physical Review, 145(1):224, 1966.

[127] Alexander I Bobenko. Geometry ii: Discrete differential geometry. Lecture Notes, Math Dept., TU Berlin, Germany, 2015.

[128] Alexander I Bobenko TU Berlin, Peter Schröder, John M Sullivan, and Günter M Ziegler. Discrete Differential Geometry, volume 38. Springer Science & Business Media, 2008.

[129] Kevin Burrage, PM Burrage, and Tianhai Tian. Numerical methods for strong solutions of stochastic differential equations: an overview. Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 460(2041):373–402, 2004.

[130] William H Press, Saul A Teukolsky, William T Vetterling, and Brian P Flannery. Numerical recipes 3rd edition: The art of scientific computing. Cambridge university press, 2007.

Appendix A

Discretization of the Helfrich Hamiltonian

Two constraints are imposed on the membrane dynamics:

- membrane closure: $\sum_{k=1}^{N} \mathbf{y}_k = 0$;

- the segments' length: $|\mathbf{y}_k| = 1$.

The first constraint is implemented by introducing a Lagrange multiplier $\mathbf{h} = (h_1, h_2)$, whereas for the second constraint N Lagrange multipliers λ_k are needed. Thus, the Hamiltonian associated to a membrane configuration becomes

$$H_{\mathrm{mem}} = V_{\mathrm{H}} + V_{\mathrm{L}} + V_{\mathrm{L}'} = \frac{\kappa_H}{2} \sum_{k=1}^{N} (C_k - C_{0,k})^2 + \mathbf{h} \cdot \sum_{k=1}^{N} \mathbf{y}_k + \sum_{k=1}^{N} \lambda_k y_k. \qquad \text{(A.1)}$$

For the Langevin equation one needs to calculate the bending forces:

$$\frac{\partial (V_H + V_L + V_{L'})}{\partial \mathbf{y}_k} = \frac{\partial V_H}{\partial \mathbf{y}_k} + \mathbf{h} + \lambda_k \widehat{\mathbf{y}}_k \tag{A.2}$$

$$0 = \frac{\partial (V_H + V_L + V_{L'})}{\partial \mathbf{y}_k} \cdot \widehat{\mathbf{y}}_k = \frac{\partial V_H}{\partial \mathbf{y}_k} \cdot \widehat{\mathbf{y}}_k + \mathbf{h} \cdot \widehat{\mathbf{y}}_k + \lambda_k = \mathbf{h} \cdot \widehat{\mathbf{y}}_k + \lambda_k$$

$$\Rightarrow \lambda_k = -(\mathbf{h} \cdot \widehat{\mathbf{y}}_k)$$

$$\frac{\partial (V_H + V_L + V_{L'})}{\partial \mathbf{y}_k} = \frac{\partial V_H}{\partial \mathbf{y}_k} + \mathbf{h} - (\mathbf{h} \cdot \widehat{\mathbf{y}}_k) \cdot \widehat{\mathbf{y}}_k$$

$$\sum_k \mathbf{y}_k = 0 \Rightarrow \sum_k \left[\frac{\partial V_{H+L+L'}}{\partial \mathbf{y}_k} \right] = 0$$

$$\sum_k (\mathbf{h} \cdot \widehat{\mathbf{y}}_k) \cdot \widehat{\mathbf{y}}_k - N\mathbf{h} = \sum_k \left[\frac{\partial V_H}{\partial \mathbf{y}_k} \right]$$

$$h_1 \frac{1}{N} \sum_{k1}^N \widehat{y}_{k,1} \widehat{\mathbf{y}}_k + h_2 \frac{1}{N} \sum_{k1}^N \widehat{y}_{k,2} \widehat{\mathbf{y}}_k - \mathbf{h} = \frac{1}{N} \sum_k \left[\frac{\partial V_H}{\partial \mathbf{y}_k} \right] = \mathbf{b}.$$

To guarantee numerical stability, two elastic-like contributions are introduced. The complete expression of the membrane Hamiltonian is then given by

$$H_{\text{mem}} = V_{\text{H}} + V_{\text{L}} + V_{\text{L}'} + V_{\text{E}} + V_{\text{E}'} =$$

$$= \frac{\kappa_H}{2} \sum_{k=1}^N (C_k - C_{0,k})^2 +$$

$$+ \mathbf{h} \cdot \sum_{k=1}^N \mathbf{y}_k + \sum_{k=1}^N \lambda_k y_k + \tag{A.3}$$

$$+ \frac{\kappa_E}{2} \sum_{k=1}^N (|\mathbf{y}_k| - 1)^2 + \frac{\kappa_{E'}}{2N} \left(\sum_{k=1}^N \mathbf{y}_k \right)^2 .$$

To calculate the bending forces for the Langevin equation (Eq. 7.2), it is convenient to express $\mathbf{y}_k$ in terms of complex numbers:

$$\mathbf{y}_k \to \omega_k = |\omega_k| e^{i\phi_k} = z_k^2. \tag{A.4}$$

Thus, the membrane dynamics is given by

$$
\begin{aligned}
\dot{\omega}_k &= -\left[\frac{\partial V_H}{\partial \mathbf{y}_k} + \mathbf{h} - (\mathbf{h} \cdot \hat{\mathbf{y}}_k) \cdot \hat{\mathbf{y}}_k + \frac{\partial V_E}{\partial \mathbf{y}_k} + \frac{\partial V_{E'}}{\partial \mathbf{y}_k} \right] + \xi \hat{\mathbf{y}}_k^{\perp} \\
&= -\left[\frac{i\omega_k}{|\omega_k|^2} \frac{\partial V_H}{\partial \phi_k} + \eta - \frac{\omega_k Re(\eta \bar{\omega}_k)}{|\omega_k|^2} + \kappa_E(|\omega_k| - 1)\frac{\omega_k}{|\omega_k|} + \frac{\kappa_{E'}}{N} \sum_k \omega_k \right]
\end{aligned}
\tag{A.5}
$$

where $\mathbf{h} = (h_1, h_2)$ in terms of complex numbers becomes $\eta = \eta_R + i\eta_I$ and

$$
\begin{aligned}
\frac{\partial V_H}{\partial \phi_k} &= \frac{\kappa_H}{2} \sum_k 2(C_k - C_{0,k}) \frac{\partial}{\partial \phi_n}\left(2\tan\frac{\phi_{k+1} - \phi_k}{2} \right) \\
&= \kappa_H \left[\frac{|z_k|^2}{|z_{k-1}|^2} \frac{\frac{\mathrm{Im}(\frac{z_k}{z_{k-1}})}{\mathrm{Re}(\frac{z_k}{z_{k-1}})} - \frac{C_{0,k-1}}{2}}{\left[Re(\frac{z_k}{z_{k-1}}) \right]^2} - \frac{|z_k|^2}{|z_{k-1}|^2} \frac{\frac{\mathrm{Im}(\frac{z_{k+1}}{z_k})}{\mathrm{Re}(\frac{z_{k+1}}{z_k})} - \frac{C_{0,k}}{2}}{\left[Re(\frac{z_{k+1}}{z_k}) \right]^2} \right]
\end{aligned}
\tag{A.6}
$$

and the corresponding equations to 7.2 are

$$
\dot{z}_k = \frac{\dot{\omega}_k}{2z_k}.
\tag{A.7}
$$

Appendix B

Vesicle extraction algorithm

In this simplified model of molecular sorting where the membrane is represented as a one-dimensional lattice, the formation of a vesicle occurs when the membrane lattice selfintersects. If that case, the newly formed vesicle is detached from the system along with the proteins it contains. In order to check intersections only between segments which happen to be close to each other, the space containing the membrane lattice is subdivided into squared cells of unitary side and the following algorithm has been implemented:

i) consider a segment with endpoints $\mathbf{n}_1$ and $\mathbf{n}_2$ (Fig. B.1);

ii) check if other lattice nodes are contained in the same cells of $\mathbf{n}_1$ and $\mathbf{n}_2$ (consider, for instance, the situation in Fig. B.1: $\mathbf{p}_1$ and $\mathbf{p}_2$ are contained in the same cell of $\mathbf{n}_1$);

iii) for each node found in the cells containing the segment $\overline{\mathbf{n}_1\mathbf{n}_2}$, we consider the incoming and outgoing segments;

iv) check intersections only between segments identified at step (iii);

v) repeat the same procedure for all membrane segments.

Checking intersection

The algorithm to check if an intersection between segments any pair of segments has occurred consists of the following steps:
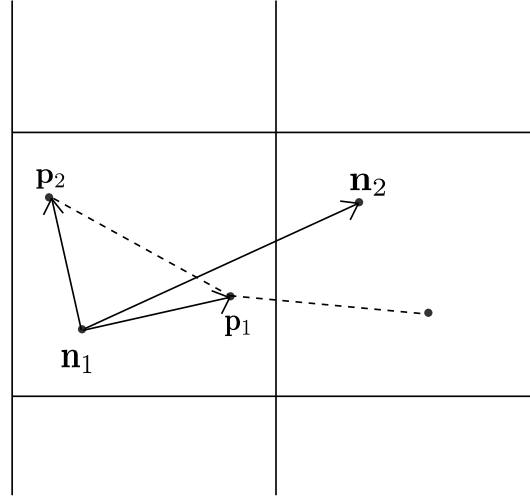
Fig. B.1 Intersections are checked only between segments of the membrane which are close with each other.

i) check if $\mathbf{p}_1$ and $\mathbf{p}_2$ are located on opposite sides with respect to the segment $\overline{\mathbf{n}_1\mathbf{n}_2}$:

$$(\mathbf{n}_2 - \mathbf{n}_1) \times (\mathbf{p}_1 - \mathbf{n}_1) \cdot (\mathbf{n}_2 - \mathbf{n}_1) \times (\mathbf{p}_2 - \mathbf{n}_1) < 0 \qquad (\text{B.1})$$

ii) If condition B.1 is verified, check the following one

$$\text{sign}(D_2)\,\text{sign}(D_4) < 0, \qquad (\text{B.2})$$

where D_2 and D_4 are given by

$$D_2 = \det \begin{vmatrix} n_{2,R} - n_{1,R} & p_{1,R} - n_{1,R} \\ n_{2,I} - n_{1,I} & p_{1,I} - n_{1,I} \end{vmatrix} = \text{Im}(UV) \qquad (\text{B.3})$$

and

$$D_4 = \det \begin{vmatrix} n_{2,R} - n_{1,R} & p_{2,R} - n_{1,R} \\ n_{2,I} - n_{1,I} & p_{2,I} - n_{1,I} \end{vmatrix} = \text{Im}(UW) \qquad (\text{B.4})$$

and, for convenience, $\mathbf{n}_1$, $\mathbf{n}_2$, $\mathbf{p}_1$, $\mathbf{p}_2$ are expressed in terms of complex numbers and we define

$$U = n_1 - n_2, \quad V = p_1 - n_1, \quad W = p_2 - n_1. \qquad (\text{B.5})$$

iii) If Eq. B.2 is verified, intersections are found by solving the following system

$$\mathbf{n}_1 + \lambda(\mathbf{n}_2 - \mathbf{n}_1) = \mathbf{c} + \mu(\mathbf{p}_2 - \mathbf{p}_1), \tag{B.6}$$

one obtains

$$\lambda = \frac{D_1}{D_3}, \quad \mu = \frac{D_2}{D_3}, \tag{B.7}$$

where

$$D_1 = \det \begin{vmatrix} p_{1,R} - n_{1,R} & p_{1,R} - p_{2,R} \\ p_{1,I} - n_{1,I} & p_{1,I} - p_{2,I} \end{vmatrix} = \mathrm{Im}(\bar{V}X) \tag{B.8}$$

$$D_3 = \det \begin{vmatrix} n_{2,R} - n_{1,R} & p_{1,R} - p_{2,R} \\ n_{2,I} - n_{1,I} & p_{1,I} - p_{2,I} \end{vmatrix} = \mathrm{Im}(UX), \tag{B.9}$$

with $X = V - W = p_1 - p_2$. The intersection is found when the following conditions are verified

$$0 \leq \frac{D_1}{D_3} \leq 1 \Rightarrow D_1 D_3 \geq 0 \Rightarrow \begin{cases} D_3 \leq 0 \Rightarrow D_1 \geq D_3 \\ D_3 \geq 0 \Rightarrow D_1 \leq D_3 \end{cases} \tag{B.10}$$

$$0 \leq \frac{D_2}{D_3} \leq 1 \Rightarrow D_2 D_3 \geq 0 \Rightarrow \begin{cases} D_3 \leq 0 \Rightarrow D_2 \geq D_3 \\ D_3 \geq 0 \Rightarrow D_2 \leq D_3 \end{cases}, \tag{B.11}$$

which implies

$$D_3 \geq 0, D_1 \geq 0, D_2 \geq 0, D_1 \leq D_3, D_1 \leq D_2 \tag{B.12}$$

or

$$D_3 \leq 0, D_1 \leq 0, D_2 \leq 0, D_1 \geq D_3, D_1 \geq D_2. \tag{B.13}$$