Self-organized computation in the far-from-equilibrium cell

Jeremy L. England Ⓘ

Check for updates
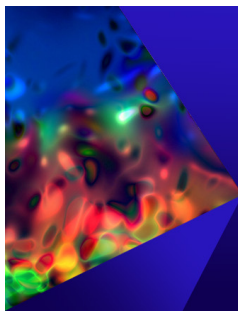
View Online

Export Citation

Self-organized computation in the far-from-equilibrium cell (F)

View Online   Export Citation   CrossMark

Jeremy L. England (iD)

AFFILIATIONS

School of Physics, Georgia Institute of Technology, 837 State St NW, Atlanta, Georgia 30332, USA
and GSK.ai, GlaxoSmithKline, 46 Menachem Begin, Ninth Floor, Tel Aviv, Israel

ABSTRACT

Recent progress in our understanding of the physics of self-organization in active matter has pointed to the possibility of spontaneous collective behaviors that effectively compute things about the patterns in the surrounding patterned environment. Here, we describe this progress and speculate about its implications for our understanding of the internal organization of the living cell.

Published under an exclusive license by AIP Publishing. https://doi.org/10.1063/5.0103151

TABLE OF CONTENTS

INTRODUCTION

Well into the 20th century, life at the cellular level was still largely an enigmatic marvel. When Erwin Schrödinger penned his famous monograph "What is life?" in 1944,[1] it may by then have seemed reasonable to suppose that the cell's whole exquisite array of chemical catalysis, transduction of molecular signals, and generation of controlled mechanical forces must somehow happen in obedience to known physical laws governing atoms and molecules—yet, still, it was nearly impossible to imagine how. In only a few decades, however, the situation would be very different, in major part due to the discovery of the so-called Central Dogma of molecular biology: Nucleic acids contained coded instructions for the assembly of proteins from amino acids, and the amino acid sequence, in turn, granted the protein a functional, self-assembled structure. Finally, a consistent intellectual thread could be traced from Darwin and the genes to Anfinsen and the structures,[2] so that a simple and compelling explanation for the development of biomolecules of diverse function could be posited: natural selection would sift DNA sequences until special ones were discovered that coded for the construction of tiny, highly useful polypeptide nanomachines. The assembly and orchestration of the magnificent pageant of molecules that made life possible started to be plausible and sensible as a physical process.

Ensuing years of further investigation have revealed a variety of ways in which this story is too neat and simple. Some polypeptide sequences do not fold into functional proteins without the help of other proteins.[3] Some traits are heritable through epigenetic effects mediated by covalent modifications of DNA or intergenerational transmission of protein aggregates.[4,5] Nonetheless, it is fair to say that whenever biologists discover another way in which living cells succeed in doing something that helps them to survive and flourish, there is a strong inclination to explain this success with some kind of Darwinian story in which the genetic sequence reigns supreme. DNA is commonly presented as the program that the cell is compiling and running, and part and parcel to this idea is the notion that the architecture of life "figures out" how to solve the problems presented by the environment by throwing out the "bad" versions of DNA coding faulty programs, and keeping the successful ones that code for better ones.[6] This paradigm certainly deserves the top-billing that it gets, for it provides the best explanation of the data in many instances. Still, a physicist looking at the cell as a collection of interacting particles obeying simple rules should be willing to ask: could there be more to the story, at least in principle?

It turns out that several streams converging from different literatures have begun to point together at a new answer to this question. In cell biology, advances in microscopy have revealed an economy of adaptive and multi-functional membrane-less structures whose feats of self-assembly may harness a variety of effects from nonequilibrium statistical mechanics.[7] At the same time, there is also a growing recognition that the collective behavior of molecules in the living cell is capable of acting as a substrate for memory and computation that can be quite functionally useful.[4,8,9] Finally, progress in our understanding of the physics of active matter suggests that a variety of spontaneous

14 September 2024 20:23:18

learning and computing behaviors may be surprisingly easy to get "for free" from the general principles governing self-organization in active mixtures exposed to patterns.[10,11] Taken together, these distinct sources of evidence and concepts provide a new perspective on biological adaptation, in which the clever herd behavior of macromolecules takes center stage and the traditionally separate activities of sensing, computation, and action become unified.

The notion that form and function are intimately connected is fundamental to biological science. Macroscopic anatomical features of a living thing may have been what first inspired this line of thinking, but it has since been applied successfully on the molecular scale. For example, many individual proteins in living cells have three-dimensional native structures that make them exceptionally good at particular biochemical tasks, such as enzymatic catalysis.[12] In contrast with the macroscale, however, the nanoscale much more directly invites us to think about the statistical aspects of the form–function relationship. There are many different ways a protein may be structured, but the full diversity of shapes that are possible in a polypeptide composed of a few thousand atoms is still more limited and primitive than the range spanned by lungs and hearts, and this inevitably raises the question of how likely one would be to get the same apparent success in function on the nanoscale "by accident."

The particular case of proteins is fairly well-studied, and while it is true that randomly assembled amino sequences have some capacity to fold and catalyze,[13] there is overwhelming evidence that the sequences of proteins coded in the genome are highly nonrandom,[14,15] presumably as a result of eons of natural selection in favor of sequences coding for proteins with exceptionally useful architecture. Zooming out to larger scales, however, there starts to be less evidence for the assumptions that get most commonly made about how good architecture arises. The living cell as a whole displays a tremendous range of form–function relationships in its spatial structure and dynamics and is clearly in a highly nonrandom arrangement of its constituent parts at any given time. Yet, we frequently encounter an explanation for this higher level structure that still traces virtually everything back to the primary protein sequence and gene regulatory elements coded by the DNA: proteins are conceived of as useful macromolecules whose native structures are a consequence of their sequences; these native structures give them specific and diverse affinities for their binding partners (such as specific nucleic acid sequences), and they are imagined to explore the cell at random through diffusion until they can happen upon the right binding partners and stick to them in the right way.[16] DNA sequence determines the shapes and affinities and catalytic capacities of the proteins and also controls when different proteins get produced in greater or lesser quantities, and in this traditional view, all form that serves function flows from the nonrandom, evolutionarily selected nature of this sequence.

The view outlined above is compellingly true to a great extent and summarizes enormous amounts of biological evidence amassed over decades. At the same time, however, we must note that it strongly emphasizes the genetic code as the explanation for all of the living cell's dynamic successes in responding and adapting to its environment. If a metabolic stress is overcome, that must be because some genes were selected to code for receptors and transcription factors. If a mechanical stimulus activates a behavior, it must be because genes coding for structural filaments have evolved to produce interactions with the right phosphorylating enzymes. Past generations of cellular life experienced many environments, and the way to find advantage in those environments got written into the expression regulation and linear sequence of proteins and now it is so well-baked that it can adapt and respond successfully to a huge diversity of novel situations. So the thinking goes.

Useful and accurate as such a picture is, however, it leans heavily on a hasty assumption. The computers that we build in factories are not useful for anything until they get meticulously programmed. Perhaps as a result, when we observe behavior and dynamic structure in living things that looks like it is computing things about its environment and acting clever about the results of those computations, we assume the bio-computer had to have been programmed beforehand. In this loose way of thinking, the DNA plays the role of the carefully constructed software. Yet, a physicist who is truly unburdened by the biases of our experience with molecular biology and human-made computers should be willing to ask the empirical question: does matter always need to be programmed in advance in order for it to compute complex and useful things about its inputs? If we have tended to answer "yes," in the past, it is because we have assumed we are only interested in getting the matter to follow a program of our own devising and compute what we want it to. Now, however, we may also ask: If a somewhat motley and random collection of interacting particles or components gets stimulated with various patterned energy sources, what does it compute by default, if we do nothing to arrange or constrain things by design?

Recent work on the selection and adaptation in active collectives has established a new picture of nonequilibrium many-body behavior that should start to make us think differently about what might be possible in the midst of the macromolecular mess. Groups of driven particles far more primitive in their motions and physical interactions than real proteins show the surprising ability to detect complex patterns and respond with specific behaviors. In what follows, we first explore these advances in physical understanding, with an eye toward what such phenomena could mean for the functioning of biological systems.

DISSIPATIVE ADAPTATION AND LOW RATTLING

Active matter is a class of nonequilibrium many-body system in which individual particles or components have access to energy from external driving forces. A variety of active matter systems ranging from protein motor and filament mixtures,[17,18] to chemically catalytic colloid suspensions[19] or robot swarms[11] have been studied experimentally, and they are known to exhibit a range of striking collective behaviors. Spiral vortices, regular crystals, and synchronized dances have all been observed in instances where many interacting active elements are subjected to nonequilibrium driving.

Active matter may provide an exciting empirical experimental playground for discovery and characterizing interesting nonequilibrium collective phenomena, but it has proven a highly challenging arena in which to make theoretical prediction. Unlike in classic self-assembly scenarios, the Boltzmann distribution linking probabilities of states to their free energies cannot be employed, and kinetic factors often reign supreme in determining the character of dynamical evolution and steady-state behavior. Of course, one way forward is to make bespoke models specific to the system of interest, and this tack has been successful many times.[20] Meanwhile, different approaches have been explored for trying to identify more universal characteristics of

14 September 2024 20:23:18

active mixtures, some focusing on the possibility of phase transitions,[7,21] and others inspired by the use of symmetry in phenomenological modeling of equilibrium condensed matter.[22] Nonetheless, the power of the Boltzmann description, in which a locally measurable property of individual microstates can be used to predict the global distribution over configurations, has proved challenging to develop.

Early on in the study of statistical mechanics away from equilibrium, the possibility that thermodynamic factors could govern the self-organization of such active systems was already appreciated. Nobel Laureate Ilya Prigogine showed that near-equilibrium driven systems reach a condition of minimal entropy production in their steady-state, which hinted at the possibility of uncovering a more general such statement for the non-linear, far-from-equilibrium regime. In anticipation of such progress—and with an eye toward understanding the architecture of living things—Prigogine famously coined the term "dissipative structures" in reference to the diversity of self-organized structures that can form in open systems driven far from equilibrium. However, while it is clear in many specific examples of such structures that the absorption and dissipation of work energy is essential for the maintenance of their orderly state, there was no general quantitative principle akin to minimal entropy production that could be shown to apply in the non-linear regime.[23,24] Thus, while the Prigoginian program offered inspiration to those searching for a link between non-equilibrium thermodynamics and biology, more progress in the theory of nonequilibrium statistical mechanics was needed before that link could be identified.

That progress came with the advent of the so-called "fluctuation theorems:" new, general results in statistical thermodynamics governing the arbitrary far-from-equilibrium and non-linear regime. Using only basic principles such as conservation of energy and time-reversal symmetry, theorists such as Chris Jarzynski and Gavin Crooks showed[25,26] how thermodynamic quantities such as work, heat, and free energy were related to the probabilities of different observations in nonequilibrium systems. Perhaps the most fundamental statement was that of Crooks, who showed that the relative probability of a dynamical trajectory $\pi(x(t))$ and its time-reversed movie $\pi^*[x(t)]$ are related simply to the heat Q released during the forward trajectory by $Q[x(t)] = k_B T \ln \left\{ \pi[x(t)] / \pi^*[x(t)] \right\}$. The Crooks relation does not imply a generalization of the minimum entropy production principle that holds in the linear, near-equilibrium regime; indeed, it makes clear instead that no such simple thermodynamic principle must govern all nonequilibrium steady-states. Subsequently, it has been shown that the Crooks relation enables a rigorous argument for an exact quantitative relationship between the stability of ordered nonequilibrium structures and the likely amounts of work absorbed during the history of their formation.[27,28]

A way to operationalize this idea that has emerged recently succeeds by separating between local and global dynamics. Nothing dynamical can be truly specific to one configuration since dynamics are all about traversing a set of different states. However, for many-body systems, there are local regions of configuration space where the small, short-time motions may have characteristics specific to that region. Specifically, it can be the case that the same external driving forces acting on a system may activate motion that is more or less orderly, or more or less violent, depending on what state the system starts in while being driven. A simple example of this is the "living crystal" experiment in which self-propelled colloidal particles form

regular crystalline arrays despite having only repulsive inter-particle forces.[19,20] When the particles are dispersed and separate, each one gets propelled from place to place. On the other hand, when the particles are jammed together into a dense array, the randomly oriented, individual propulsion forces acting on them average out to zero and the nonequilibrium driving produces very little motion.

The experimental example described above turns out to be a simple and intuitive instance of a more general phenomenon known as low rattling. Low-rattling behavior was first identified through the observation that the same nonequilibrium system can be capable of exhibiting randomly diffusive exploration of configuration space, orderly motion, or stasis, just depending on what part of configuration space it happens to be in while being exposed to the same external drive.[10,11] This effect has now been established in a variety of simulated and experimental systems, ranging from disordered spin glasses and mechanical networks to robot swarms. The basic principle is that a choice of external drive (such as a time-varying force) will determine the amplitude and degree of diffusive randomness in the local motion of each configurational state available to some active matter system of interest. After a long time, the dynamics are predicted to bias the system toward dwelling in states where diffusive motion (rattling) is weak, either due to the onset of dynamical order, or due to the overall reduction in drive energy converted into motion within the system. To put the intuition in more primitive terms: complex driven systems gravitate toward states from which it is hard for the external drive to eject them.

There is a loose analogy to machine learning here; a neural network is a complicated function whose many parameters determine a specific way that input quantities are mapped to output quantities, and the algorithmic "training" of a neural network selects these parameters so that the input–output mapping represented optimizes a chosen figure of merit involving the network's output agreeing with target values. The key point with low rattling, by comparison, is that it is a physical mechanism that effectively imposes a specific figure of merit on the behavior of a driven system so that its dynamics will gradually adapt the many internal variables describing the system configuration to reduce the rattling of the output given the pattern of drive input. Thus, a complex many-body system exhibiting low rattling may be thought of as a peculiar example of a machine learner whose optimization of input/output did not need to be programmed algorithmically because it gets implemented for free by the physical dynamics.

Self-organization by low rattling can be expected to occur in any partly chaotic, damped, driven many-body system for which various parts of configuration space exhibit different dynamical response properties. In simulated mechanical networks, this has been shown to allow for adaptive modulation of the spectrum of resonance frequencies under oscillatory driving,[29] and in experimental studies of robot swarms, the same principle has been shown to allow for the selection of desired collective dynamical states.[11] Moreover, in the example most akin to computation and machine learning, a simulation of a network of randomly coupled discrete binary variables subject to time-varying random external fields exhibited an emergent novelty-detection ability in which the energy absorption of the network was observed in simulation to spike precipitously when the pattern of external driving changed from an initial set of random driving fields to a new set to which the system was naïve.[30] This example suggests a powerfully expressive capability to learn complex patterns in the

timing of environmental force fluctuations. In this regard, low rattling serves as one class of a broader range of possible dissipative adaptation behaviors, in which it has been argued that the externally driven exploration of a many-body configuration spaces will show bias toward final states with exceptionally dissipative histories of work absorption.[27,28] There is no known rigorous relationship between the near-equilibrium minimum entropy product principle of Prigogine and the far-from-equilibrium low-rattling effect which can also lead to reduction in dissipation. However, low rattling does offer a practically applicable optimization principle for far-from-equilibrium self-organization of the sort that Prigogine's program hoped to uncover.

The low-rattling picture of collective behavior in active matter contrasts with more classic experiments in the field. Many earlier studies of nonequilibrium self-organization have focused on how organized complexity can emerge unexpectedly when many-body systems are driven in simple ways at the single-unit level.[17,18,20] Low-rattling ideas are still in early days, but their key lesson is clear: by driving with patterns of forcing that are richly structured, it is possible to study a new kind of exceptionality in the emergent organization that has to do with what is being computed about or adapted to in the statistics of the drive. There may be many amenable experimental systems where such paradigm has yet to be explored effectively.

LIFE FAR FROM EQUILIBRIUM

There has likely never been a point since the full development of statistical mechanics as a rigorous theory when someone who understood it well would have said that living things exist at thermal and chemical equilibrium. The reasons are legion: organisms are open systems that act as conduits for fluxes of energy and matter that enter and leave in distinct forms. These flows, in turn, are essential for sustaining matter in otherwise unstable states that are constantly rolling downhill in free energy. Life also exhibits at all scales a variety of behaviors and phenomena that have clear directionality in time, violating the famous equilibrium principle of detailed balance that forbids clock-like cycles. Suffice it to say, living things are some of the most strikingly nonequilibrium processes we know.

Nonetheless, much of the history of applying physicochemical ideas to biology is written in the language of equilibrium. There are many good reasons why equilibrium concepts come in handy when making sense of the physics of life. Sometimes, it is the case that the interaction between two macromolecules under dilute, isolated conditions *in vitro* can reveal a great deal about what underlines their functioning together in the cell, and sometimes proteins fold and unfold reversibly in a test tube. Thus, both because of limitations to early methods of observation and measurement, and also because many aspects of molecular biophysics can quite successfully be studied by isolating components to study their equilibrium properties, the concepts of thermal and chemical equilibrium permeate the way we talk about the physical processes that make living cells tick. The folding of proteins is often talked about as a case of equilibrium self-assembly where the native tertiary fold is favored by the minimization of free energy.[12] The binding and unbinding of ligands to their molecular partners is described with the language of affinity and dissociation constants. Even the power strokes of very much nonequilibrium molecular motors get their energy bookkeeping in a language that refers to standard states of chemical fuels.[16] Of course, motors are just one example of a wide variety of processes that are highly

nonequilibrium overall, but whose dynamics can, nonetheless, be described in terms of detailed balance-breaking chemical currents that transit from one well-defined local equilibrium state to another. Therefore, in one sense, the attempt to break life apart into a catalogue of equilibrium binding affinities and Gibbs free energies is understandable. Nonetheless, the example of protein folding shows the slippery slope here: while it is sometimes temptingly convenient to think of a protein's tertiary structure as a matter determined by the equilibrium properties of a polypeptide chain on its own, the fact is that many proteins do not fold properly without interacting with ATPase chaperones and other nonequilibrium processes. Examples like these point to the need to tread carefully and recognize that the early development of our understanding of molecular and cellular biology skewed toward what early methods could reveal; more recent work has cast a new light on the inner workings of the cell that emphasizes instances where biological structure and function demand nonequilibrium language from the outset.

Cells, when construed as collections of atoms, are far from equilibrium in at least three crucial respects that ultimately can be unified in one physical framework. First, they are far from chemical equilibrium, meaning that the composition of different types and amounts of different molecules does not reflect the simple prediction of a Boltzmann distribution governed by Gibbs free energies.[16] Second, they are in conformational disequilibrium, since many macromolecules persist and function in various kinds of kinetically trapped shapes.[31] Finally, they are in locational disequilibrium, since the macromolecules that makeup most the cell's dry weight are arranged across space and with respect to each other in ways that are tightly influenced by driven transport processes.[21,32]
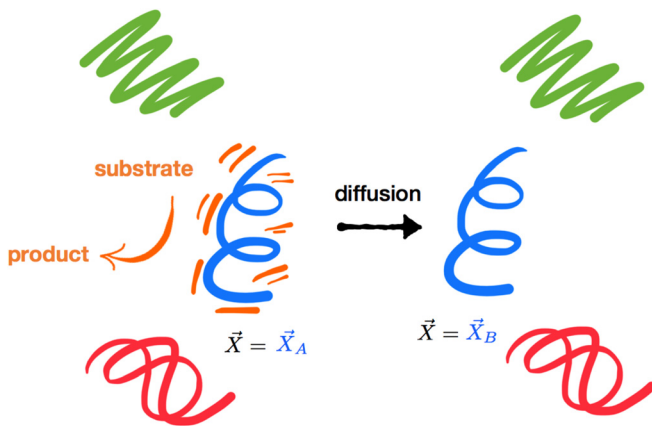
Before discussing each of these instances in more detail, it is worth noting that they all look the same to the eye of a certain kind of statistical physicist. Leaving the added (and, in this case, irrelevant) complexities of quantum mechanics aside, the "full" microscopic description of the physical state of a cell can be thought of as the detailed specification of all the positions and velocities of all nuclei and electrons out of which the cell is built. Once a coordinate system is chosen for such a fine-grained description, chemical, conformational, and locational aspects of the physical state basically reduce to different scales of description. Whether we are talking about reactions that form and break chemical bonds, folding and unfolding events that alter the tertiary configuration of a protein, or the larger scale question of which proteins and metabolites are concentrated in which parts of the cell, all of these transitions can be thought of as changes in the same coordinates specifying the overall structural state.

In a true equilibrium, the probability of observing the system in a fully specified coordinate state X would simply be $\exp(-\beta E(X))/Z$. In a living thing, the state of affairs is far from satisfying such an expression, for many reasons. Peptide bonds, for example, are at higher Gibbs free energy than hydrolyzed amino acids and would be unlikely to appear at physiological temperature and pressure. The kinetically trapped peptide bonds of cellular protein, however, are constantly being regenerated by new protein synthesis and take a long time to fall apart spontaneously once formed.[33] The truth is that countless degrees of freedom in the molecular state of the cell are similar in quality to the peptide bond: pumped up into metastable states of high free energy that would never be visited in a Boltzmann distribution and strongly reflect the far-from-equilibrium bias in system state
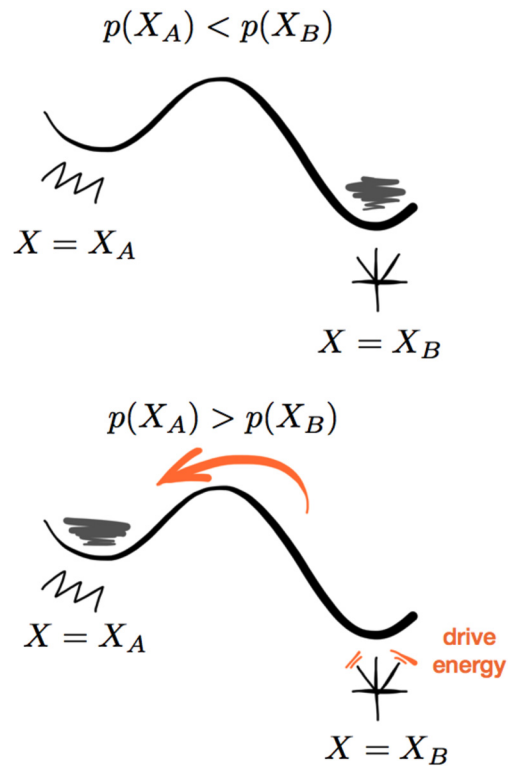
that obtains due to the constant external forcing. This bias manifests simply in $p(X)$ deviating from Boltzmann form for certain subsets of elements of X.

The same kind of bias also can manifest in the values of X that reflect the conformation of a protein (such as whether it is natively folded or in a denatured, unfolded "coil state") or also where in the cell the protein is located and how it is oriented in whatever oligomeric complexes it joins. Though the physical factors that influence the Boltzmann distribution via $E(X)$ still matter far from equilibrium, since they refer to forces that push things together or apart under any condition, it remains the case that many coordinates of the cell have to be thought of us being far from Boltzmann in their probability. Although the Anfinsenian paradigm[12] of reversible folding of globular protein domains contributed greatly to our understanding of macromolecular structure, it is, nonetheless, the case that a broad variety of abundant and important proteins in bacteria and eukaryotes require the help of ATPase molecular chaperones to get folded into the right functional shape.[31] Similarly, although many proteins exhibit diffusion and explore the confines of the cell seemingly at random, the effective diffusion constant with which they do so may depend both on their interactions with other proteins,[32] and even on their own enzymatic activity[34,35] (Fig. 1). In practice, this means that whether we are talking about the shape, locations and orientations, or chemical integrity of all proteins in the cell, it is fair to talk about the state of affairs in terms of a damped, driven many-body system described by a coordinate X whose distribution of states is determined by the deviations from Boltzmann behavior that the particular pattern of driving and kinetic trapping brings about (Fig. 2).
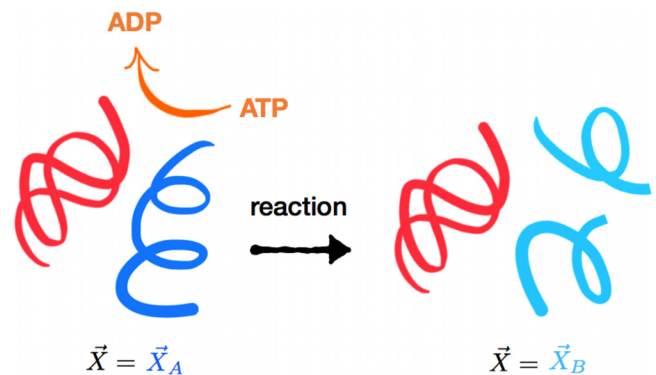
The most basic and best understood way in which cells are not at equilibrium has to do with the chemical fluxes that sustain the populations of different chemical species at levels inconsistent with the standard free energies (Fig. 3). In an animal cell, the directional flow that sustains this state of affairs is metabolism, which constantly takes input such as sugars and uses their stepwise breakdown to power the production of widely usable chemical fuels such as nucleoside triphosphates like ATP and GTP. We often focus on ATP and GTP especially

FIG. 2. Two assembly states described by different coordinate values X_A and X_B are separated by a barrier and differ in energy (which increases in the upwards vertical direction). In thermal equilibrium (above), the Boltzmann distribution dictates that probability density should accumulate more at X_B, which has lower energy. However, in the presence of nonequilibrium driving (orange), it is possible for external forcing to pump probability density over the barrier, such that it accumulates in the higher energy state. One possible mechanism for this might be if the assembled structure corresponding to X_B is better at absorbing work energy from the patterned external drive than the assembled structure described by X_A.

FIG. 1. Some changes in the spatial coordinates describing the molecules in a cell may involve transport of a macromolecule from one location to another. For example, the coordinate vector $\vec{X}$ might describe the center of mass of a protein (blue) whose own enzymatic activity might cause it to diffuse more rapidly than other non-catalytic proteins of comparable size and shape (red and green).
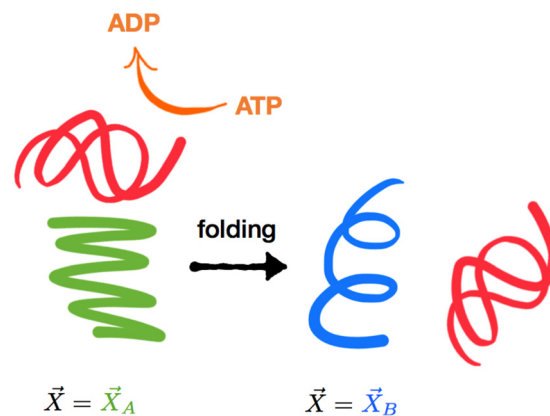

FIG. 3. Some changes in the spatial coordinates describing the molecules in a cell may involve chemical transformations. For example, the coordinate vector $\vec{X}$ might describe the atomic degrees of freedom of a protein (blue) whose hydrolysis might be catalyzed by a protease (red) either with or without active chemical driving (orange).

because the free energy gradient they are sustained in becomes a "water wheel" that many other cellular processes can hook into, but the truth is that chemical disequilibrium is rampant across the cell's whole composition. Protein, RNA, and DNA makeup most of the cellular mass aside from water. Both in order to build free-energy-rich bonded structures anabolically, and in order to pay for the irreversibility that guarantees long-term kinetic stability despite their thermodynamic instability,[27,33] the formation of these macromolecules consumes nucleoside triphosphates (NTPs) and deoxyribose nucleotide triphosphates (dNTPs) at a prodigious rate. Many subsequent covalent modifications of macromolecules, the phosphorylation of proteins by kinases being the most famous, exploits the nonequilibrium population balance of chemical species for the purposes of sensing and signaling. Of course, many passive enzymes catalyze chemical reactions that would otherwise be slow, and thereby accelerate the relaxation of certain reactants and products to local equilibria.[36] However, since the overall network of metabolites is tilted by the detailed-balance-breaking flows that either produce or make use of ATP, it is never safe to assume a given metabolite is in global equilibrium. The cell as a whole is constantly relaxing toward equilibrium, and a constant influx of new free energy pays to continually kick things back uphill as various chemical species roll slowly or rapidly down.[37]

In the scene described above, there would be the risk of everything sounding relatively simple and uninteresting: a caricature of chemicals at different concentrations than expected, or where some of them participating in directional cycles of conversion and regeneration, is the one we have from freshman biochemistry. The truth is that already in chemical space,[8,9] there are a variety of interesting computational behaviors that can emerge, but in the case of biology, it is easier to see how such effects can arise by thinking not about chemical bonds, but about the conformations and locations of macromolecules. A single human cell contains tens of millions of proteins, and each one is a polymer typically composed of hundreds of amino acids. For some proteins, there is evidence that frequent cycles of reversible folding and unfolding characterize their conformational dynamics, so that the stability of the functional tertiary structure of the protein is undergirded by being favored at equilibrium as the lowest free energy ensemble of states.[12] The last several decades of have revealed, however, that a plurality of protein gene products, many of them essential for cellular life, cannot fold properly without assistance from molecular chaperones.[31] Chaperones, such as GroEL in bacteria and CCT in eukaryotes, are ATPase proteins that couple their driven conformational changes to binding and unbinding of other protein substrates (Fig. 4). The world of chaperones is diverse, and not all of them have clear physical mechanisms, but it is understood, in general, that they act as conformational "proofreaders," using ATP hydrolysis to power the reshaping of other proteins in their functional forms.

Understanding how and why chaperones work is a world in itself, but for our discussion here, their mention should suffice to point out that many proteins should be thought of as being stuck in kinetic conformational traps that persist for long periods of time if a power-assist from a molecular chaperone is not provided. Evidence of this kind of conformational trapping has cropped up in various settings, for example, in the study of so-called static heterogeneity in the activity of single enzymes.[38] Using fluorescent reporters, it is possible for certain enzymes to quantify the rate of catalysis of different copies of the protein over time, and remarkable, there is evidence that individual

FIG. 4. Some changes in the spatial coordinates describing the molecules in a cell may involve structural transformations. For example, the coordinate vector $\vec{X}$ might describe the configurational degrees of freedom of a protein with a misfolded (green) and natively folded (blue) conformational states whose native folding might be catalyzed by a chaperone protein (red), usually with the assistance active chemical driving (orange).

enzymes with the same amino acid sequence can persistently differ in the rate at which they catalyze the same reaction. The idea of persistent, kinetically trapped differences in macromolecular structure is all the more significant at the level of amyloid aggregates, a multi-protein ordered alignment of many polypeptide chains that can survive boiling in high concentrations of denaturant once formed.[39]

The point for our purposes here is that each individual protein should be thought of as having a vast space of possible conformations it can adopt, with a smaller (but still potentially vast) subset of them being metastable kinetic traps that persist on biologically relevant timescales in the presence of thermal fluctuations. Once we add in the combinatorial possibilities of multiple proteins combining to form different kinds of multi-component complexes (including, but not limited to, amyloids), the same state of affairs only intensifies. The prevalence of kinetic trapping also underlines that there is no imaginable scenario in which these complex biomaterials could ever be expected to reach equilibrium, even if they were not being chemically driven: the rugged glassiness of the energy landscape they explore ensures the system is always in some slowly relaxing out-of-equilibrium state due to starting in a non-Boltzmann initial condition. Proteins, therefore, possess an unexplorably large space of possible configurations that they only can traverse with the active, fuel-consuming help of molecular chaperones and other macromolecular processes that couple enzymatic activity to conformational change. Not only this, but the paradigm of chaperoning implies the possibility that the particular conformational changes that are stimulated, and the particular proteins or protein complexes that come to interact with chaperones, can be selected with specificity based on the affinities between particular proteins due to their amino acid sequence.

The cellular population of macromolecules is also in a thoroughly nonequilibrium state when it comes to the spatial locations of all the different cytosolic and other components. The most famous reason for this is the existence of so-called active transport mechanisms, which exploit highways of filaments that serve as conduits for ATP-driven motors that can drag loads processively from one place to another.

However, more recent work has pointed to an even greater diversity of reasons that external driving impacts transport. There is evidence that some proteins diffuse more or less freely in the cell as one expects from Brownian motion. However, for others, molecular chaperones rear their head again in this case because of how they impact solubility and the effective size of a typical macromolecular complex in which a protein finds itself; indeed, it can be safely assumed that without the active help of ATPase chaperones, the diffusion constant of the cytosol as measured by fluorescence recovery after photobleaching (FRAP) would drop precipitously due to massive and irreversible misfolding and aggregation.[32] Consequently, the very possibility of diffusive transport must be thought of as partly a product of active driving that can be modulated by access to protein folding quality control machinery.

More exotically, recent experiments have also shown that even a protein enzyme that engages in passive catalysis of an exothermic reaction may experience a substantial kick from the accompanying release of energy, such that the enzyme itself has an effective diffusion constant that is affected by its access to substrate in its local vicinity.[34] The idea that access to substrate might impact diffusive motion is an intriguing one given evidence and theoretical thinking that supports possibility that enzymes that participate in catalysis of sequential reactions may colocalize to achieve efficiencies.[40]

Indeed, in recent years, a whole literature has emerged around increasing indications that cells nucleate membrane-free multi-protein assemblages by exploiting the physics of liquid–liquid phase separation. Phosphorylation of linker proteins[21,41] can modulate the tendency of large numbers of liquid/gel components to condense in one locale, and this condensation, in turn, can form the basis for creating a microenvironment with altered concentrations of other proteins. The formation of such membrane-free bodies in the cell has been implicated in a variety of stress-response behaviors, even to the point of it being suggested that each one may be a unique new creation and not a programmed reproduction of a set composition or structure.[42] There also has been evidence of such stochastic cluster formation being an important event in eukaryotic transcription initiation. There has been a marked tendency in attempts to understand these bodies to draw an analogy to equilibrium phase separation, following the example of oil droplets in water. In equilibrium, however, the only thing that should affect how co-localized protein components tend to be should be their binding affinities and their bulk concentrations. In this new, nonequilibrium world, each protein's local access to enzymatic substrates, cycling covalent modifications, or ATPase binding partners, such as motors or chaperones, all could impact its tendency to accumulate and participate in a local condensate. The upshot is that a much richer palette of possibility is available in the self-assembly of these intracellular bodies, in ways that have the opportunity to reflect the details of how energy sources and molecular components fluctuate dynamically in time.

NEW OPPORTUNITIES

Taking these various observations together, it becomes possible to think of the cellular proteome as a whole as a fruitful ground for low rattling and other dissipative adaptation behaviors. On the one hand, the particular set of conformations and bound complexes the cell traverses at a particular moment is going to be determined by the way that active processes such as chaperoning move proteins between different conformational states. One the other hand, the overall conformational state of the cell at any given time also clearly impacts how

much chemical fuel in the cell gets converted into conformational changes by such a mechanism, since chaperones interact with their substrates in a conformation-specific way. This implies a feedback loop, whereby access to chaperoning and other active mechanisms both determines the future conformations and spatial distributions of proteins and also thereby determines their future access to additional active input. One thus expects a biased search in a vast space of possible configurations that should settle on a dynamical attractor in which active processes are less able to disrupt and transform the distribution of protein conformations.[10]

In particular, when it comes to chaperone activity, a simple version of the above argument has been known for a while as iterative annealing.[43] Iterative annealing was long ago proposed as a possible mechanism of action for GroEL, whereby it would preferentially bind to and mechanically unfold the more hydrophobic conformations of its substrate, thus ensuring that the protein in question would get as many chances to find the hydrophilic native state as it needed and not get stuck in hydrophobic and misfolded kinetic traps. The generalization of this idea through the lens of low rattling simply points out that any conformational feature of the proteome as a whole that alters the ability of ATP to power further conformational change can lead to a selection effect on the long-term conformational features that are observed. Since the relationship between conformation and access to binding partners and the energy they may deliver can be highly subtle and depending on specific protein–protein binding interactions, and since active driving can bring about more diverse changes than unfolding, this opens up the possibility that far more complex patterns than bulk hydrophobicity might be sensed, integrated, and processed into predictive behavior by a large active collective such as the cytosol.

This last point deserves careful unpacking, because it risks being mistaken for a less provocative statement. It is quite easy to accept the idea that the living cell could turn out to carry out computation-like activities relevant to successful biological function using all sorts of intracellular components. Aside from the fact that efforts in synthetic biology have demonstrated cellular components can be turned into mathematical calculators with the right engineer,[8,9] we take for granted anyway that information processing is something cells have long-since been shaped by evolution to do in myriad ways. What the low-rattling idea brings into our perspective on cytosol is the notion that a heterogeneous active collective in a fluctuating environment might tend to act as a complex pattern detector even before any Darwinian selection were brought to bear on how such a computation could be used by the cell to aid survival and reproduction at the organismal level. To put it more bluntly, even if all proteins had randomly distributed affinities for each other, and if the way these affinities were modified by ATP-powered conformational change were also random, it would still be the case that low rattling would be expected to act as an optimization principal governing the input–output relation of the collective organization achieved by the active mixture. Individual macromolecules would experience pressure over time to arrange themselves over space and in such a distribution of conformational states so that whatever patterned temporal or spatial variations in access to ATP were determined by the environment, the randomizing diffusive motion in the system would eventually "learn" to decrease by moving more in step with those predictable fluctuations. This sort of machine-learning-like activity is something we would expect from such an active collective without having to program it or, in particular, engineer the interactions of its components.

There would be an upper limit to how interesting it could be for our understanding of biology to know that large active collectives of biomolecules might compute complex things about their physical and chemical environments if such a process were always an accidental sideshow to the real functioning of life. However, while it is important to establish that the self-organized computational behaviors implied by low rattling are something that can be gotten for free, as it were, there is little reason to think that living organisms undergoing Darwinian evolution would pass up the opportunity to exploit such a capability if it existed in their cells. In a scenario where amino acid sequences are subject to the replicative iterations of mutation and selection in a Darwinian scenario, it is easy to envision how the particular signaling events that are coupled to particular conformational changes, and the particular protein–protein interactions that lead to spatial co-localizations, could easily couple the outputs of "accidental" computations performed by the collective to machinery that uses them in more obvious ways to store or communicate sensed information. The way to think about implications of low-rattling self-organization for biology, then, is to appreciate that life may have a potentially useful means of computation built into its molecular structure, and we always should be asking the question of how and whether this means might turn out to be part of the explanation for functional success we observe in a biological context.

There is much that is suggestive in past experiments and simulations in non-biological settings, but the question now remains how to use experimentation in the future to test whether the potential described here gets realized in ways that impact biology. One might envision at least two approaches. The first would be to let the biological phenomena of interest drive the discussion and go in search of processes essential to the function of living things that are still in need of elucidation. The study of membraneless inclusion bodies seems ripe for further investigation in this regard, since the diversity and specificity with which different granules and droplets seem capable of condensing with functionality targeted at the needs of the particular event of cell stress is suggestive of a very flexible adaptive capability.[42]

The notion of spontaneous pattern-detection or computation in the sub-cellular herd behavior of macromolecules should justifiably be viewed with skepticism when regarded as a mere speculation about what might be possible in principle given what we know about physics. Nonetheless, there are known examples already of computing-like behaviors implemented by the molecules of the cell. The yeast prion Sup35, for example, was shown to act as a heritable form of memory storage.[4] Sup35 forms prionic aggregates that are capable of catalyzing duplication of their structural form. Sup35 aggregates interact with the Hsp104 ATPase chaperone and can propagate heritably from one generation to the next in a way that depends on the chaperone's activity. In addition, however, the presence or absence of Sup35 impacts the cellular phenotype, becoming a non-genetic mechanism for storing memories of the environment experienced by previous generations. From a physical perspective, this example hints strongly at the importance of kinetically trapped macromolecular structures and their interactions with sources of active, ATP-powered conformational remodeling. This non-genetic mechanism of interaction between environment and structural proteome is known now to be a common means of inheritance of phenotype in wild yeasts.

Gene regulatory networks present another example of how the far-from-equilibrium high-dimensionality of the biological system can become the substrate for richly expressive spontaneous computational behaviors. The cellular levels of different gene products vary over time, with each gene having the potential to directly or indirectly influence the future expression of itself or another gene. Complex many-body dynamical systems have previously been identified as being capable of acting as "reservoirs" for embodied computing, even in macroscopic soft robotics settings.[44] In the case of gene regulatory networks, even more specific claims have been studied in simulation, in work that points to the potential for Pavlovian conditioning of associative memories. The high-dimensional dynamical space of the gene expression network is too vast to explore fully, such that the current state of network always can hysteretically reflect the system's previous trajectory and interaction with its environment. Using Boolean Kauffman models to simulate regulatory network dynamics, Levin lab has shown that even randomly wired gene regulatory networks can learn to associate two co-applied stimuli (A and B) with each other so that eventually a behavior caused by stimulus A can be evoked subsequently by exposure only to stimulus B.[45] A tantalizingly similar scenario has already been described in a recent study of promiscuous receptor ligand interactions in the bone morphogenic signaling pathway,[46] in which a collective of many components has been shown to integrate and process information for complex decision-making.

At the same time, it might be argued there is an inherent challenge in beginning with the biological phenomenon if the goal is to find the smoking gun of adaptive computation in the cellular setting. There are always many factors that contribute to the system level behavior of a collection of macromolecules, and establishing that the outcome is the result of a spontaneous computation by the cell as opposed to a "hard-coded" response written into the protein sequences by more conventional evolutionary processes is bound to be challenging. There also is, therefore, great need for experimentation building from the other side of the gulf, in which biomolecular components are used to construct a new generation of active matter experiments that push the pattern-learning aspect to the fore. A new literature of nonequilibrium physics experiments has begun to emerge that provides a valuable model for this approach, in settings as diverse as photonic materials[47] and colloidal acoustic media.[48] The fundamental question always has to be: can I demonstrate in my system an emergent, fine-tuned response property that is matched to the complex pattern of an external drive input? The more this model of experimental design can be tailored to active matter settings involving *in vitro* protein mixtures or cytosol extract, with arbitrary patterns set by the experimenter used as the environmental challenge, the more it will be possible to demonstrate the principle of this type of computation in biological material. Success in this area would no doubt set the stage for a smarter approach to looking for instances where the cell clearly already uses this kind of phenomenon to its advantage.

ACKNOWLEDGMENTS

AUTHOR DECLARATIONS
Conflict of Interest

The author has no conflicts to disclose.

14 September 2024 20:23:18

Author Contributions

Jeremy L. England: Conceptualization (equal); Investigation (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

REFERENCES

[1]E. Schrodinger, *What is Life* (Cambridge University Press, 1944).

[2]P. A. Romero and F. H. Arnold, "Exploring protein fitness landscapes by directed evolution," Nat. Rev. Mol. Cell Biol. **10**, 866–876 (2009).

[3]M. R. Leroux and F. Ulrich Hartl, "Protein folding: Versatility of the cytosolic chaperonin TRiC/CCT," Curr. Biol. **10**, R260–R264 (2000).

[4]J. Shorter and S. Lindquist, "Prions as adaptive conduits of memory and inheritance," Nat. Rev. Genet. **6**, 435–450 (2005).

[5]T. A. Chernova, Y. O. Chernoff, and K. D. Wilkinson, "Prion-based memory of heat stress in yeast," Prion **11**, 151–161 (2017).

[6]S. Kryazhimskiy, D. P. Rice, E. R. Jerison, and M. M. Desai, "Global epistasis makes adaptation predictable despite sequence-level stochasticity," Science **344**, 1519–1522 (2014).

[7]Y. Shin, Y.-C. Chang, D. S. Lee, J. Berry, D. W. Sanders, P. Ronceray, N. S. Wingreen, M. Haataja, and C. P. Brangwynne, "Liquid nuclear condensates mechanically sense and restructure the genome," Cell **175**, 1481–1491 (2018).

[8]R. Daniel, J. R. Rubens, R. Sarpeshkar, and T. K. Lu, "Synthetic analog computation in living cells," Nature **497**, 619–623 (2013).

[9]O. Purcell and T. K. Lu, "Synthetic analog and digital circuits for cellular computation and memory," Curr. Opin. Biotechnol. **29**, 146–155 (2014).

[10]P. Chvykov and J. England, "Least-rattling feedback from strong time-scale separation," Phys. Rev. E **97**, 032115 (2018).

[11]P. Chvykov, T. A. Berrueta, A. Vardhan, W. Savoie, A. Samland, T. D. Murphey, K. Wiesenfeld, D. I. Goldman, and J. L. England, "Low rattling: A predictive principle for self-organization in active collectives," Science **371**, 90–95 (2021).

[12]C. I. Branden and J. Tooze, *Introduction to Protein Structure* (Garland Science, 2012).

[13]K. M. Digianantonio, M. Korolev, and M. H. Hecht, "A non-natural protein rescues cells deleted for a key enzyme in central metabolism," ACS Synth. Biol. **6**, 694–700 (2017).

[14]N. Halabi, O. Rivoire, S. Leibler, and R. Ranganathan, "Protein sectors: Evolutionary units of three-dimensional structure," Cell **138**, 774–786 (2009).

[15]D. S. Marks, L. J. Colwell, R. Sheridan, T. A. Hopf, A. Pagnani, R. Zecchina, and C. Sander, "Protein 3D structure computed from evolutionary sequence variation," PLoS One **6**, e28766 (2011).

[16]S. L. Kittsley, *Physical Chemistry Principles and Applications in Biological Sciences*, edited by K. Sauer, I. Tinoco, Jr., J. Puglisi, and J. C. Wang (Pearson, 1979).

[17]V. Schaller, C. Weber, C. Semmrich, E. Frey, and A. R. Bausch, "Polar patterns of driven filaments," Nature **467**, 73–77 (2010).

[18]T. Sanchez, D. T. Chen, S. J. DeCamp, M. Heymann, and Z. Dogic, "Spontaneous motion in hierarchically assembled active matter," Nature **491**, 431–434 (2012).

[19]J. Palacci, S. Sacanna, A. P. Steinberg, D. J. Pine, and P. M. Chaikin, "Living crystals of light-activated colloidal surfers," Science **339**, 936–940 (2013).

[20]G. S. Redner, M. F. Hagan, and A. Baskaran, "Structure and dynamics of a phase-separating active colloidal fluid," Phys. Rev. Lett. **110**, 055701 (2013).

[21]P. Li, S. Banjade, H.-C. Cheng, S. Kim, B. Chen, L. Guo, M. Llaguno, J. V. Hollingsworth, D. S. King, S. F. Banani *et al.*, "Phase transitions in the assembly of multivalent signalling proteins," Nature **483**, 336–340 (2012).

[22]E. Putzig and A. Baskaran, "Phase separation and emergent structures in an active nematic fluid," Phys. Rev. E **90**, 042304 (2014).

[23]I. Prigogine, "Time, structure, and fluctuations," Science **201**, 777–785 (1978).

[24]D. C. Spanner, "Biological systems and the principle of minimum entropy production," Nature **172**, 1094–1095 (1953).

[25]C. Jarzynski, "Nonequilibrium equality for free energy differences," Phys. Rev. Lett. **78**, 2690 (1997).

[26]G. E. Crooks, "Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences," Phys Rev E **60**, 2721–2726 (1999).

[27]J. L. England, "Dissipative adaptation in driven self-assembly," Nat. Nanotechnol. **10**, 919–923 (2015).

[28]N. Perunov, R. A. Marsland, and J. L. England, "Statistical physics of adaptation," Phys. Rev. X **6**, 021036 (2016).

[29]T. Kachman, J. A. Owen, and J. L. England, "Self-organized resonance during search of a diverse chemical space," Phys. Rev. Lett. **119**, 038001 (2017).

[30]W. Zhong, J. M. Gold, S. Marzen, J. L. England, and N. Yunger Halpern, "Machine learning outperforms thermodynamics in measuring how well a many-body system learns a drive," Sci. Rep. **11**, 9333 (2021).

[31]F. U. Hartl, A. Bracher, and M. Hayer-Hartl, "Molecular chaperones in protein folding and proteostasis," Nature **475**, 324–332 (2011).

[32]R. Gura Sadovsky, S. Brielle, D. Kaganovich, and J. L. England, "Measurement of rapid protein diffusion in the cytoplasm by photo-converted intensity profile expansion," Cell Rep. **18**, 2795–2806 (2017).

[33]J. L. England, "Statistical physics of self-replication," J. Chem. Phys. **139**, 121923 (2013).

[34]C. Riedel, R. Gabizon, C. A. Wilson, K. Hamadani, K. Tsekouras, S. Marqusee, S. Pressé, and C. Bustamante, "The heat released during catalytic turnover enhances the diffusion of an enzyme," Nature **517**, 227–230 (2015).

[35]S. Pressé, "A thermodynamic perspective on enhanced enzyme diffusion," Proc. Natl. Acad. Sci. **117**, 32189–32191 (2020).

[36]J. R. Knowles and W. J. Albery, "Perfection in enzyme catalysis: The energetics of triosephosphate isomerase," Acc. Chem. Res. **10**, 105–111 (1977).

[37]J. M. Horowitz, K. Zhou, and J. L. England, "Minimum energetic cost to maintain a target nonequilibrium state," Phys. Rev. E **95**, 042102 (2017).

[38]X. Michalet, S. Weiss, and M. Jäger, "Single-molecule fluorescence studies of protein folding and conformational dynamics," Chem. Rev. **106**, 1785–1813 (2006).

[39]T. P. Knowles, M. Vendruscolo, and C. M. Dobson, "The amyloid state and its association with protein misfolding diseases," Nat. Rev. Mol. Cell Biol. **15**, 384–396 (2014).

[40]M. Castellana, M. Z. Wilson, Y. Xu, P. Joshi, I. M. Cristea, J. D. Rabinowitz, Z. Gitai, and N. S. Wingreen, "Enzyme clustering accelerates processing of intermediates through metabolic channeling," Nat. Biotechnol. **32**, 1011–1018 (2014).

[41]Y. Shin and C. P. Brangwynne, "Liquid phase condensation in cell physiology and disease," Science **357**, eaaf4382 (2017).

[42]D. Kaganovich, "There is an inclusion for that: Material properties of protein granules provide a platform for building diverse cellular functions," Trends Biochem. Sci. **42**, 765–776 (2017).

[43]D. Thirumalai, G. H. Lorimer, and C. Hyeon, "Iterative annealing mechanism explains the functions of the GroEL and RNA chaperones," Protein Sci. **29**, 360–377 (2020).

[44]K. Nakajima, T. Li, H. Hauser, and R. Pfeifer, "Exploiting short-term memory in soft body dynamics as a computational resource," J. R. Soc. Interface **11**, 20140437 (2014).

[45]S. Biswas, S. Manicka, E. Hoel, and M. Levin, "Gene regulatory networks exhibit several kinds of memory: Quantification of memory in biological and random transcriptional networks," iScience **24**, 102131 (2021).

[46]Y. E. Antebi, J. M. Linton, H. Klumpe, B. Bintu, M. Gong, C. Su, R. McCardell, and M. B. Elowitz, "Combinatorial signal perception in the BMP pathway," Cell **170**, 1184–1196 (2017).

[47]C. Ropp, N. Bachelard, D. Barth, Y. Wang, and X. Zhang, "Dissipative self-organization in optical space," Nat. Photonics **12**, 739–743 (2018).

[48]N. Bachelard, C. Ropp, M. Dubois, R. Zhao, Y. Wang, and X. Zhang, "Emergence of an enslaved phononic bandgap in a non-equilibrium pseudo-crystal," Nat. Mater. **16**, 808–813 (2017).

14 September 2024 20:23:18