Dynamics and Evolution of Modular, Hierarchical Structures in Natural and Engineered Biological Systems

1. Results from Prior NSF Support

Award PHY-0943390: Collaborative Research: Synthetic Integrons for Continuous Directed Evolution of Complex Genetic Ensembles. PI: Joshua Leonard; $473,293 (9/1/09-8/31/13) launched 6/1/10. Collaboration via separate grant to PI: Jay Keasling (UC Berkeley). This project was a 6-university international collaboration that grew out of an NSF/EPSRC "Sandpit," which targeted high-risk synthetic biology projects (1). Overall objectives: construct a technology enabling continuous directed evolution of multi-gene functions using novel biochemical and computational tools and optimize a biosynthetic pathway. **Intellectual Merit:** Developed experimental and computational tools for tracking dynamic evolution of genetic heterogeneity in multi-plasmid systems and technology for controlling gene flow in microbial networks. **Broader Impacts:** trained 2 graduate (3 years funding) and 1 undergraduate student; collaboration trained 5 postdoctoral fellows and 3 graduate students. Given the profile and international scope of this project, this work enhanced many outreach and training activities detailed in Dr. Leonard's Biosketch**.** This work also enabled a second phase of this multinational collaborative project, which was recently funded. This new work will leverage the experimental and computational tools developed in the first phase of the project to build a comprehensive toolkit for the synthetic biology community, enabling the rapid generation of biosynthetic industrially and medically-relevant products. **Publications** include: reviews of challenges and opportunities in synthetic biology (2, 3) and experimental and computational technologies for probing and controlling the dynamics of genetic elements within engineered microbial populations (4, 5).

Award PHY-0848755: Experimental and Theoretical Analysis of Collective Dynamics in Swarming Systems. Subaward PI: Cristián Huepe; $194,100 (9/1/09-8/31/13) (main-award PI: I D Couzin, Princeton U; $543,472). This project studied swarms experimentally and theoretically from a non-equilibrium statistical physics perspective. **Intellectual Merit:** Experiments (Couzin): implemented fully controlled fish swarming experiments and tracking. Theory (Huepe): developed detailed and idealized models to capture specific and universal aspects of swarm dynamics using experimental results and theory; introduced adaptive-network description of collective motion; extended standard swarming models to include variable speeds, visual interactions, etc.; characterized transition dynamics between states of collective motion; developed new approaches for understanding collective decision making in animal groups. **Broader Impacts:** Co-trained 4 graduate students & 4 postdoctoral researchers. Organized interdisciplinary meetings: *Arts, Humanities and Complex Networks* (NetSci2012); *Networks and Nonlinearities in the Musical Experience* (ZiF Ctr for Interdiscip Rsrch, Germany). Public lectures and roundtable on *Physics of Music* at Chicago Cultural Ctr., Scientific consulting for *eduMedia* educational website. Developed control algorithms for swarm robotics engineering. Media coverage: Wired Magazine, Natl. Geo., Focus Online, etc**.** (Main award PI, I.D. Couzin, leads CouzinLab & teaches at Princeton U, participates in advisory panels, often gives public lectures & media interviews). **Publications** (coauthored by C. Huepe) include work on: collective motion in active elastic membranes and with variable speeds (6, 7), network-based analyses of the adaptive voter model (8), experimental studies of collective behavior in schooling fish (9) (10), swarm robotics control applications (11, 12), adaptive network models of swarm dynamics (13), and interdisciplinary music/science research (14); plus over 25 other publications by ID Couzin et al. – 2 in Science and 5 in PNAS. **Other products and activities:** Data, models and code available at CouzinLab's and Huepe's websites. Book chapter on 'Flocking and Music' for *Controls and Art* (Springer). Member of Advanced Study Group *Statistical Physics of Collective Motion* (Max Planck Inst., 2011-2013).

Dr. Dirk Brockmann has not been a recipient of prior NSF awards.

2. Introduction

Transformative approaches for characterizing living systems have produced substantial quantitative information describing biological processes at the molecular, cellular, organismal and even population levels. Translating these data into *understanding* is a grand challenge and opportunity for this century (15). Much of this information can be effectively represented by various biological interaction networks, including genetic regulatory networks, protein interaction networks, metabolic networks on a microscopic scale, interaction networks of cells in tissues, neural networks on a mesoscopic scale, and the interaction of species on a large scale in ecological networks. Biological interaction networks thus exist across many scales and generically exhibit a fractal-like architecture, in which basic building blocks are combined to

form more complex structures, which are in turn organized to form structures at yet higher levels. This *modular* and *hierarchical* (MH) organization impacts dynamic properties of these systems, since interactions between building blocks are also organized as a hierarchy of processes and subprocesses, often spanning multiple timescales. Here, we refer to systems that follow such organization of either interaction topologies or dynamics as MH *systems* or MH *structures*. Figure 1 schematically depicts an MH structure. Only three levels are displayed, but the illustrated hierarchy can be repeated any number of times to form a multi-scale MH structure.

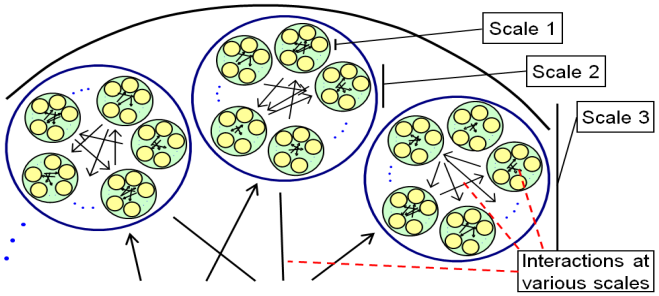

Figure 1: Schematic representation of modular hierarchical structure. Interacting modules organize in a hierarchy of levels or scales. Small yellow discs are modules that interact (through arrows) to yield the dynamics of the embedding green discs (modules at the next scale). These interact in turn within the embedding blue modules to produce dynamics at the next interaction scale.

Despite the ubiquity of MH structures in living systems, we still lack formal tools for defining and characterizing them, and have limited understanding of their effects on system dynamics and on evolutionarily-important properties such as robustness and fitness. The purpose of this proposal is to study and understand MH systems from an integrated theoretical and experimental viewpoint by combining approaches from complex networks, non-equilibrium dynamics, statistical physics and synthetic biology. We will be guided by the following central, overarching and open questions:

- **What types of dynamics are supported by MH structures?**
- **Do MH structures improve fitness or confer other evolutionary advantages?**
- **How do MH structures emerge and evolve?**

The existence of MH structures in biological networks is evidenced by a number of recent studies in which data from different biological systems was analyzed to reveal underlying organizational principles. These studies have identified at least three topological and dynamical features that evidence MH structures, namely, the presence of *motifs*, *modules*, and a *hierarchical organization*. For example, genetic regulatory networks contain a small set of recurring patterns, called *network motifs*, with similar regulation functions (16). Specific motifs, classified into families such as feedforward loops, single-input modules, or dense overlapping regulons, have been identified in *E. coli* (17, 18), yeast (19, 20), and other organisms (21-23). Motifs can be viewed as the lower-level organizational units of a MH structure, with interactions among motifs forming the next hierarchical level. The presence of underlying MH architectures is also supported by the studies of *modularity* (24) in biological systems, which emerged in parallel in evolutionary developmental biology (25, 26) and molecular systems biology (27, 28). This abstract property is observed in a variety of contexts and is believed to be an essential part of biological organization, related to the many levels and types of functional and structural heterogeneity in living systems (29, 30). While specific formal definitions vary, a well-established working definition describes a module as a part of an organism that integrates a set of processes and operates comparatively autonomously with respect to other modules in a system (24, 31). Biological networks with modular organization include protein-protein interactions in yeast (32, 33), gene regulatory networks in *Drosophila* and in processes of higher organisms such as vertebrate myogenesis (34), co-varying traits of evolving organisms associated to modular morphogenesis (35, 36), and RNA secondary structure (37). As in the case of motifs, modularity implies at least two structural levels: one for the internal functioning of the modules and the other for interactions between modules. A natural extension of these analyses is to consider subsequent levels of organization in which structures of structures are identified. This corresponds precisely to the recently considered property of having a *hierarchical* organization. While motifs and modules require only a two-level hierarchy, it is harder to identify multi-level hierarchies. Several approaches for extracting these structural properties from the architecture of natural networks have been developed (38-42). While the unambiguous identification of modules at different levels of the hierarchy is still an open problem, there is growing consensus that these are recurrent structures in biological networks (43, 44). Indeed, topologies displaying hierarchical modularity have been identified in a variety of biological networks (45, 46). However, the relationship between topological modules defined

based on connectivity and functional units remains unclear, and thus it is not known whether these structures emerge due to natural selection or rather due to other underlying processes (47).

From a theoretical perspective, other recent studies have considered the question of which network architectures are compatible with the MH structures observed in nature. When modular topology is combined with the approximately power-law, long-tail *degree distribution* (number of links per node (48)) also common in biological networks, an apparent contradiction emerges. Indeed, a characteristic feature of scale-free networks is the presence of a small number of nodes with many interactions. These highly connected nodes, or hubs, tend to integrate the whole system in a way that would seem incompatible with a naive picture of relatively autonomous modules. However, in a study of 43 distinct organisms (49) it was shown that their metabolic networks are organized into a hierarchy of highly connected modules that are combined into successively larger and less connected structures, satisfying simultaneously the conditions of modularity and scale-free connectivity. The resulting architecture, defined as a *hierarchical network*, also reproduces other features of metabolic networks, such as their clustering coefficient (50). Various deterministic hierarchical algorithms for constructing networks that are both modular and scale-free have been introduced (49, 51-53). These serve as a starting point for the investigations in **Section 4.2** below.

The studies described above focus on the network topology, without addressing the evolutionary dynamics that is implied (either as a cause or as a consequence) by it. Evolution is a strong unifying principle in biology and an essential component for the development of living systems at all scales, from basic genetic processes to complex interactions between species. In spite of its apparent universality and simple conceptual basis, the process through which random fluctuations and natural selection lead to the remarkable level of self-organization observed in living systems is far from understood. Notably, the role and mechanisms of natural selection at different levels of a MH structure is still the subject of intense debate (54, 55). Many models have been developed to try to reproduce the essential properties of evolving systems. These include a variety of approaches such as the analysis of minimal cellular automata (56) and Boolean network models (57), the use of concepts from game theory (58) and graph theory (59), and the implementation of Genetic Algorithms (60) and Artificial Life programs (61) like *PolyWorld* (62), *Tierra* (63), *Avida* (64), *ECHO* (65), and *AntFarm* (66), to name a few. While interesting lessons have been learned from these studies, they are not designed to explore the evolution of MH structures and lack the necessary degrees of freedom or imposed environmental pressures to sustain them. One of the objectives of the proposed project (detailed in **Aim 3** below) is to combine algorithms that generate MH networks with simulations of evolutionary dynamics, in order to study the emergence and stability of MH structures. These efforts will be motivated by recent work suggesting that modular structures emerge in systems that evolve under alternating fitness criteria (67-69).

As described above, to date, research involving MH structures has consisted of isolated approaches that include: characterizing quantitatively the level of hierarchy and modularity in real biological networks, developing network growth models that produce MH topologies, and studying examples of the dynamics supported by them. But because these investigations are disparate in both methodology and context, they do not yet enable a direct investigation of the questions that are the focus of this proposal. A key obstacle in the research of MH systems is the difficulty of validating models and theoretically derived hypotheses with experiments. Although theoretical studies are typically motivated and gauged against empirical data, novel insights cannot be tested if experiments are not specifically designed to understand MH structures and dynamics. We will follow here a different strategy and propose a research agenda where experiments and theory are tightly coupled, by engineering synthetic biology experiments that will allow us test theory adaptively and to address the same questions in parallel on the properties, causes, and consequences of MH systems in abstract models and in the real-world physical and biological context. Specifically, we will develop an experimental setup based on bacterial toxin-antitoxin (TA) systems (**Section 4.1**); modular genetic elements that influence biological processes at multiple levels of organization (70, 71).

3. Project Overview

The project aims to advance our understanding of MH structures by investigating the dynamical consequences and evolutionary origins of this ubiquitous feature of living systems. To address the fundamental questions listed in the **Introduction**, we propose to pursue an integrated research program linking recent theoretical advances with novel experimental systems that enable direct manipulation of the structural, dynamical, and evolutionary aspects of MH biological systems. Our goals are: to characterize MH systems by formalizing existing definitions and linking them to our models and experiments, to explore MH dynamics on both abstract network models and on experimental engineered MH systems,

and to investigate whether evolutionary forces favor MH structures and the effect of these structures on fitness. In particular, we will investigate: (i) the emergence of MH structures in network-based models and synthetic biology experiments; (ii) the connection between MH interaction topologies and MH dynamics; (iii) properties of dynamical processes supported by MH network models and experimental systems; (iii) hybrid models that combine network-like and diffusive couplings to describe the physical constraints of the interactions at different levels of the hierarchy; (iv) network evolution and growth algorithms that produce MH architectures and dynamics, and their extension to artificial life simulations; (v) the robustness and evolvability of dynamics on MH networks, including experimental MH systems subjected to changing environmental conditions and selective pressures. Finally, to enhance the broader impact of this work, we will develop a simple toy model that illustrates MH structures using components whose only phenotype is motion in physical space. This model provides a mechanical analogy for MH biological systems that will serve as both a tool for teaching nonequilibrium dynamics, self-organization and evolution, and as a means to engage a broad audience through an online game-based crowdsourcing experiment. Our project will follow the complementary Key Approaches listed below.

4. Key Approaches

We introduce here the family of experimental systems and class of theoretical models over which we will develop our investigations. In order to start with concrete examples of MH systems, we first describe the toxin-antitoxin processes used to engineer MH structures in our synthetic biology experiments. We then present the class of abstract network models that will be used to study the general dynamical and evolutionary features of MH systems. Finally, we describe the Network State Diffusion model, which bridges our synthetic biology experiments and its abstract network representations by taking into account the physical properties of the inter- and intra-cellular levels of the interaction hierarchy, and uses these physical constraints to clearly distinguish between the different hierarchical levels.

4.1. Experimental construction and interrogation of MH networks. An ideal experimental platform for investigating MH networks in biological systems should have the following properties: (1) it enables the construction of families of biological systems characterized by similar MH networks whose key properties (e.g., topology) can be adjusted via defined experimental manipulations in order to facilitate “apples-to-apples” comparisons linking experimental and theoretical work; (2) network states can be monitored dynamically using established and robust experimental methods; (3) it facilitates the investigation of questions of biological and/or biotechnological significance. For each of these reasons, we propose to develop an experimental platform based upon bacterial toxin-antitoxin (TA) systems (Fig. 2a).

TA systems are modular genetic elements that influence biological processes at multiple levels of organization (70, 71). For example, the canonical Type II TA systems comprise a protein toxin and protein antitoxin, where the antitoxin is typically less stable than the toxin. When encoded on a plasmid, these systems enforce plasmid maintenance, because a daughter cell that lacks the plasmid will be killed when the antitoxin pool degrades enabling the latent pool of toxin to exert its effect (72). Interestingly, many TA systems are integrated into bacterial genomes, where their role is more controversial and might involve programmed cell death (altruistic sacrifice) or the induction of dormancy when cells are exposed to stress (70, 73), which may play a role in the emergence of antibiotic-resistant “persister” strains (74). Other experiments have suggested that endogenous chromosomal TA systems confer no specific selective advantage under the conditions investigated (75). TA systems are also of industrial relevance as tools for biotechnology (76) and as targets for development of novel antibacterial therapies (77). Most importantly for our purposes, the modular, well-characterized nature of TA interactions enables us to apply the tools of synthetic biology to engineer a robust experimental platform comprising customized intracellular and intercellular networks. Synthetic biology is an emerging discipline seeking to enable the robust engineering of biological systems and is largely inspired by the proposition that novel biological functions may be engineered by recombining fundamental “parts” (e.g., promoters, genes, transcriptional regulators, etc.) into novel configurations that create gene circuits performing specified functions (2, 78, 79). This approach has been used to engineer physical model systems, including synthetic predator-prey relationships (80), programmed spatial organization of multicellular networks (81, 82), oscillators in individual cells (83, 84), coupled oscillators distributed throughout a multicellular network (85), and Turing pattern generators that have been designed and are under development (86). Here, we propose to develop new TA-based networks for testing specific questions and hypotheses of direct relevance to the overarching aims of this proposal.

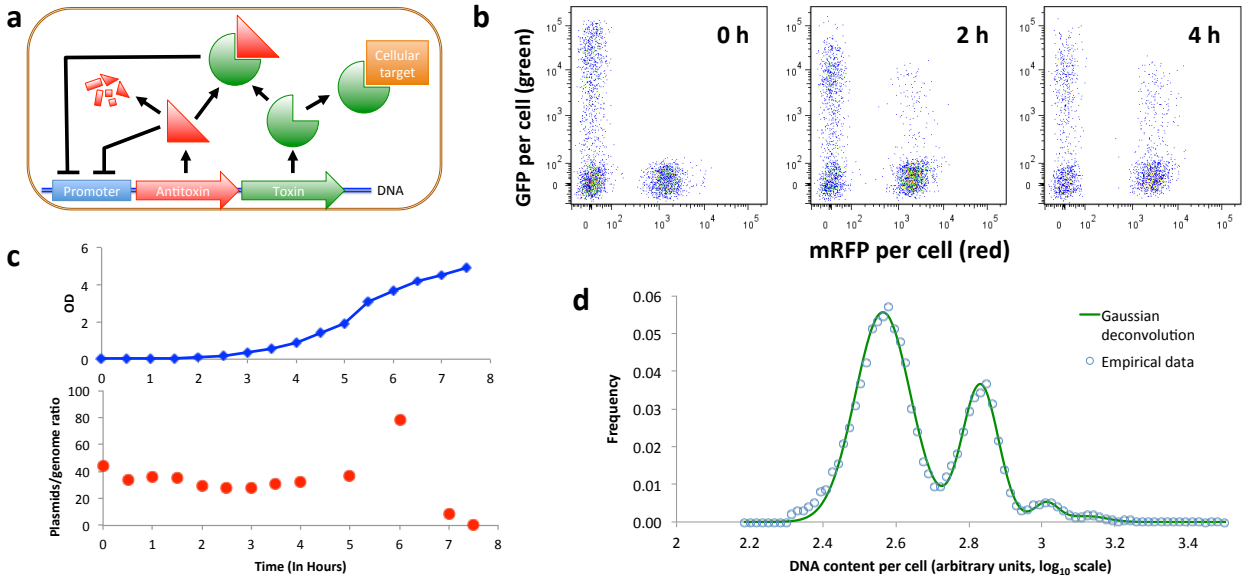

Figure 2: Experimental framework for investigating MH networks in biological systems. (a) Our model systems are based upon bacterial toxin-antitoxin (TA) networks. This cartoon summarizes the general architecture of natural type II TA systems, in which TA components are co-expressed from a single promoter, antitoxin inhibits toxin-mediated interference with cellular processes, antitoxin and toxin-antitoxin complexes negatively regulate expression of TA components, and antitoxin is degraded rapidly (e.g., by host proteases) relative to the toxin. (b) Representative single-cell data describing population dynamics, in this case illustrating transfer of a GFP-encoding plasmid (green) from donor cells to recipient cells expressing mRFP (red) over 4 hours of conjugative transfer. (c) Plasmid content varies dramatically as a function of cell growth phase (quantified by qPCR), and (d) genomic DNA content varies significantly between individual cells in a population (quantified by flow cytometry).

TA systems also provide an ideal model system for investigating a general question of biological and biotechnological relevance: how should one construct a biological network to best achieve a biological objective in manner that is stable and robust to varying environmental conditions and perturbations? For example, when engineering metabolic pathways for industrial biosynthesis, it is often necessary to balance the expression of metabolic enzymes in order to avoid accumulation of toxic intermediates or inefficient use of resources. In one demonstration, yields of a desired metabolite product were maximized over different growth conditions by engineering artificial regulatory networks that balance the concentrations of key metabolic enzymes (87). In the case of TA systems, cell growth is maximized by balancing T and A components to avoid either toxicity or squandering of cellular resources. Thus, the TA model system serves as an extreme case representing a general biological challenge. We anticipate that understanding how MH networks could stabilize TA systems would inform the design of stable engineered biological networks, which remains an open challenge in biotechnology.

In prior work, we have developed robust methodologies for quantitatively investigating the dynamics of genetic information in bacterial networks. This work establishes the key methodologies required to conduct the investigations proposed here. For example, various genetic “parts” (e.g. genes or regulators of gene expression) can be encoded on extra-chromosomal DNA elements called plasmids, and by engineering different plasmids to drive the expression of distinct fluorescent proteins, one can quantify the plasmid content of individual living cells within a population. We have used such an approach to quantify the transfer of plasmids from one cell strain to another via natural or engineered conjugative transfer systems (Fig. 2b) (5). In systematic investigations of the connection between growth dynamics and genetic content, we paired these single-cell metrics with genetic tools such as quantitative real-time PCR (qPCR) to determine that copy numbers of plasmids and even genomes vary widely between individual cells in a population (Fig. 2c,d). We have integrated these results into a novel modeling framework for characterizing and predicting such dynamics (4). Importantly for the proposed investigations, experimental perturbations that lead to heterogeneity of plasmids (or genomes) encoding TA components are expected to act as stressors on such cells, so our prior observations provide a menu of experimental perturbations that can be used to investigate the “performance” of various MH networks encoding TA components, as described in **Section 5** below.

4.2. Theoretical MH Framework. We introduce here the approach that will be used in our theoretical investigations. It is based on a family of network models that provide a common framework to study the dynamical consequences and evolutionary causes of MH structures in living systems. MH structures can be implemented in networks by imposing MH topology in their architecture or MH dynamics in the interactions between node states. The origin and evolution of MH structures can be addressed by considering network growth algorithms or link selection processes. Their dynamical response can be studied by implementing dynamical rules for the node states, which can range from continuous variables, where processes are defined as a set of ODE's per node, to Boolean variables, where each node is essentially a logical gate. The family of network models described below combines evolving network topologies with the dynamics of node states given by processes on the network. It will allow us to consider both components in the same context and using the same computational tools, to either implement a model subset for a specific investigation or to include all parts in larger simulations that thus become similar to artificial life studies of evolving systems.

MH network topologies can be generated by various recently proposed algorithms. These can be categorized as: duplication based, subdivision based, and adaptive-network based. Duplication based algorithms grow the network by making copies of the current network at every step and wiring these to produce a larger network. This process is repeated until the desired size is reached (49, 51-53, 88). Subdivision based algorithms, instead, repeatedly replace individual nodes of the current network by sub-networks, following a given protocol (89-91). Adaptive-network based algorithms evolve the connectivity based on the dynamics of node states supported by the network or on emergent properties of the topology (67-69, 92-97). Since one of our objectives is to find biological mechanisms that could lead to MH networks, we will focus here on duplication and adaptive-network based algorithms (98-100), which have clear evolutionary interpretation and can be designed to match our experimental systems. We will then explore the connection between MH topologies and MH dynamics on our resulting networks.

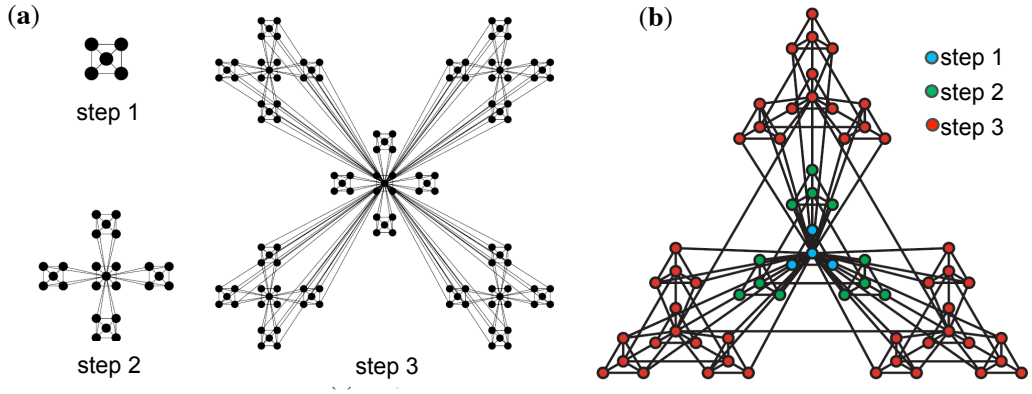

Figure 3. Examples of published duplication-based network generating algorithms (49, 52). Generation steps are displayed separately (a) or labeled by color (b). Central node connectivity makes the central module the highest hierarchical level in both cases. Direct same-level interactions between modules are included in (b) but not in (a).

MH dynamics will be studied by implementing ODE-based or Boolean-based dynamics on the generated networks. ODE-based studies will provide a closer match to experimental system, tracking in detail changes in the amounts of specific metabolites. Boolean-based dynamics allow the systematic analysis of all idealized biological dynamics supported by a given network structure, an approach first developed in the context of genetic networks (28, 101, 102). One of the simplest and best studied Boolean network models is the N-K model, or Kauffman model (103). In it, each node can only be in a 0 or 1 state and receives exactly K inputs from other, randomly selected nodes. The next state of each node is computed as a function of the state of the K nodes connected as its inputs, according to a truth-table generated at random for each node. A rich behavior emerges from this model (104). The resulting trajectories in the phase space can be stationary, periodic, or chaotic. It was argued that living organisms must be at the critical interface between both phases to be robust enough to maintain performance under a broad range of random perturbations and also evolvable enough to adapt over time (105-110). Many extension of the N-K model have been considered (111, 112). We will take advantage of the vast literature on this model and of its approach for exploring all supported dynamics in phase space to analyze the processes supported by different MH topologies. We will benefit in this effort from interactions with external (unfunded) collaborator Prof. M. Aldana, an expert in the field (see letter attached).

In order to structure our exploration of the dynamics and evolution of different MH network systems, we will consider a family of models based on a single Generalized Algorithm (GA). We start from two duplication based algorithms found in the literature (see Fig. 3). These create first a network of 5 or 4 all-to-all connected nodes (step 1). This module is then copied 4 or 3 times (step 2) and the outer nodes of the new copies connected to the central node of the original module. In panel (b) the central nodes of the new structures are also linked. These steps are repeated until the desired network size is reached. If we interpret these algorithms as representing idealized developmental and evolutionary dynamics, it is natural to extend them by introducing mutations in the module duplication steps. Further evolutionary aspects will be included using adaptive-network algorithms that modify connectivity, for which we will benefit from the expertise of external (unfunded) collaborator Prof. T. Gross (see letter attached). The analogy with evolution can be completed by adding selective pressures to the fitness of the resulting structures, defined through the emergent properties of either the topology or the resulting dynamics of network states (considering these as a proxy for the system's function). Our GA is defined as follows:

1. **Seed topology and dynamics:** Create seed module with N_m nodes connected according to an internal coupling (adjacency) matrix $\mathbf{W}_=$. Define a dynamical system on the network by specifying node states and node processes.
2. **Multiply:** Make multiple versions of the current structure either by duplicating and mutating it or by generating an ensemble of variations of it, creating a set of modules of modules after the 2nd iteration. Define the new modules as part of a lower hierarchical level (to connect accordingly below).
3. **Connect:** Define node subsets; S_-, $S_=$, S_+; that mediate each module's connectivity to others in lower, equal, or higher hierarchical levels, respectively. Connect new (lower-hierarchy) modules among them following $\mathbf{W}_=$ (now for modules rather than nodes), linking only $S_=$ nodes, and to old (higher-hierarchy) module following a $\mathbf{W}_+$ matrix, linking only lower-level S_+ nodes to higher-level S_- ones.
4. **Select:** Select nodes, links, or structures based on fitness criteria imposed on the dynamical activity of the states of specific nodes and modules. In some setups, prune the network by also discarding nodes or links that are irrelevant for the resulting dynamics of module states.
5. **Repeat:** Go to Step 2 to continue evolving and/or adding new levels to the MH structure.

Examples (a) and (b) in Fig. 3 are special cases of this GA with $N_m = 5$ and $N_m = 4$, respectively, where we skip Step 4, $\mathbf{W}_=$ and $\mathbf{W}_+$ describe all-to-all coupling, $S_=$ and S_- contain only the central node, and S_+ contains all outermost nodes of a module. Note that we will generalize these central and outer node definitions in terms of their connectivity (instead of their location on the diagrams).

This GA will provide a common framework for our investigations and also allow us to develop a single computer code with modular subroutines to carry out our studies. In some of its versions, the GA is simply a network generating algorithm that produces different MH structures for our research, in other versions it can be viewed as a simple Artificial Life system that allows us to simulate MH evolution.

4.3. Network State Diffusion model. In order to establish a direct connection between our experiments in **Section 4.1** and abstract network models in **4.2**, we define a dynamical system that takes into account the physical constraint of having mainly diffusive interactions at the inter-cellular level of the interaction hierarchy. We refer to this system as the *network state diffusion* (NSD) model. It was recently developed in the Brockmann lab and is based on the idea of combining complex network dynamical systems with reaction-diffusion dynamics. Each of these reflects a different scale or hierarchical level in the MH structures and experimental systems under investigation, here given by a spatially distributed system of interacting cells (Fig. 4). At the intercellular level, these are coupled by diffusive signaling. At the intracellular level, cell phenotype selection, differentiation and the dynamics of cell states result from the regulatory dynamics of interacting genes. These processes can be modeled using high-dimensional network-based dynamical systems where the degrees of freedom describe the gene expression levels, which interact as a set of dynamic variables through an underlying network of regulatory interactions.

The dynamical system for each independent cell takes the form

$$\partial_t \mathbf{u} = f(\mathbf{u}; \mathbf{W}) \qquad (1)$$

where elements W_{ij} of coupling matrix $\mathbf{W}$ denote interaction strengths between genes. Gene expression is denoted by the elements of the vector $\mathbf{u}$. A generic choice for the nonlinear dynamical system can be derived from underlying Michaelis-Menten kinetics, which yields Hill-type functions, for example

$$\partial_t u_i = -\gamma u_i + \sum W_{ij} \frac{u_j^\beta}{\theta_j^\beta + u_j^\beta} \quad (2)$$

This type of approach for understanding cell fate and state has been applied in many single cell models that, for example, address how the stability of sequential differentiation cascades depends on topological features of the underlying regulatory network. On a the scale of cell populations, simplified reaction diffusion models are typically employed in which the concentration of signaling molecules correspond to the dynamical quantities that interact in simplified, low-dimensional phenomenological dynamical systems of signaling, which propagate spatially by diffusion. These models can account for tissue differentiation in simplified contexts but cannot account for each cell's internal response dynamics to external signaling profiles. NSD models combine both approaches and can be cast into the form

$$\partial_t \mathbf{u} = f(\mathbf{u}; \mathbf{W}) + \mathbf{D}\, \partial_x^2 \mathbf{u} \quad (3)$$

where matrix $\mathbf{D}$ is assumed to be a sparse matrix of diffusion coefficients with non-zero elements corresponding to gene signals between cells. NSD systems can generate a rich variety of dynamical phenomena. For example, if the individual cell components possess multiple stable solutions, these stable network states can "diffuse" along the spatial coordinates and potentially trigger a propagating wave of state changes across a population of cells. NSD systems correspond directly to 2 level MH structures where local modules are spatial replicates of the intracellular coupling of interacting genes, and cross-module couplings are the diffusive links between cells mediated by signaling nodes.

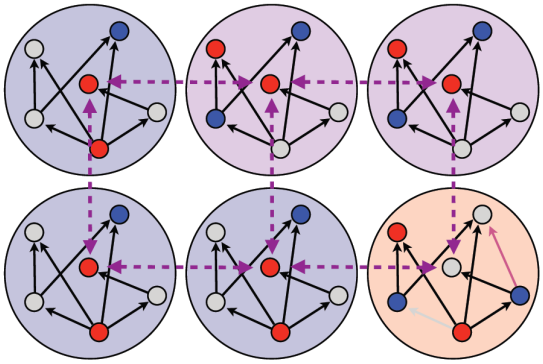

Figure 4: Example of network state diffusion model system; spatial cell lattice of internal network of interacting genes (black arrows). All cells in the array, except for the cell in the lower right, are "clones" (they possess the same network). Each network can generate a variety of stable activation states, as illustrated by node colors in the figure. Cells interact through diffusive signaling that effectively couples networks at the next hierarchical level (purple, dashed lines). Cell states thus diffuse across interfaces, thus allowing the generation of spatial patterns.

The advantage of NSD models is that they properly represent the distinct physical mechanisms that mediate interactions at the different levels of the hierarchy, corresponding here to the intracellular and intercellular levels. This will allow us to set up experiments and simulations that test key questions about MH systems, such as the robustness of internal network $\mathbf{W}$ as well as the possible dynamical outcomes of cell state interactions at the population level. Another interest of NSD models is that Eq. (3) can also be used to address evolutionary dynamics on MH systems. Genetic heterogeneity and thus phenotypic variability as a function of position can be modeled by introducing an explicit dependence of the regulatory networks on the spatial coordinates, using $\mathbf{W} = \mathbf{W}(\mathbf{x})$. Evolutionary dynamics can thus be modeled by considering slow topological modulations of $\mathbf{W}$, driven by selection rules on states generated by the entire dynamical system.

5. Integrated Experimental and Theoretical Research Plan

We will build on the experimental and theoretical approaches described above to study, in parallel, generic and specific features of the MH systems under investigation. We will examine their origins, dynamic behavior, and biologically relevant functions through an integrated research plan divided into three overarching aims, each addressing a set of specific questions through complementary experimental and theoretical investigations. Aim 1 is to generate and characterize MH structures, Aim 2 to explore their dynamics, and Aim 3 to examine how they evolve and adapt over time.

Aim 1. Develop framework for generating, classifying, and characterizing MH structures

We will first develop and characterize libraries of both experimental and theoretically-generated MH interaction networks. These will also be quantitatively characterized in order to develop metrics and classification schemes that formalize our analyses to facilitate the pursuit of subsequent aims. This

foundational work will include the development of approaches for explicitly linking experimental and theoretical studies. Key questions to be addressed include:

- How detailed should the abstract networks and NSD models be to capture the main dynamical features of modules in the synthetic biology experiments? How unique are these representations?
- What duplication-based algorithms produce MH networks with topologies similar to those found in natural and engineered (model) biological networks? How sensitive are those algorithms to noise?
- How do variations in the topologies of experimental and theoretical MH interaction networks impact their function (i.e. fitness) with respect to a given set of performance metrics?
- Do different MH topologies exhibit different levels of robustness to perturbations?

Aim 1.1 Generate an experimental library of model MH systems. As described in **Section 4.1**, our experimental framework will comprise engineered microbial networks based on TA systems (Fig. 2a). In this subaim, we will use a set of tools from synthetic biology to generate a library of interaction networks with various features and topologies that display different levels of modular and hierarchical organization. These systems will include:

Individual TA networks within a given cell: We will engineer various plasmid-based gene circuits exhibiting topologies like those depicted in Figure 2a. Circuits will utilize one of two well-characterized Type II TA systems from *E. coli.* The first TA pair is based upon *mazF-mazE*, in which the toxin (MazF) is a ribosome-independent mRNA interferase. The second TA pair is based upon *relE-relB*, in which the toxin (RelE) is a ribosome-dependent mRNA interferase (71). In each case, if toxin is present at levels higher than can be sequestered by the antitoxin, new protein synthesis is blocked, leading to cell dormancy and death. Genomic copies of the *mazF*, *maze, relE,* and *relB* components will be deleted to prevent interference with engineered components, a modification that should not alter host physiology (75). Various plasmids will be constructed, introducing variations that include: (a) either Maz or Rel-based components, (b) different plasmid copy numbers, and (c) changes in the relative translation efficiency of toxin and antitoxin components (modulating the ribosome binding site in front of T or A components will alter the resulting stoichiometry; predictive tools exist for making such modifications (113)).

Coupled TA networks within a given cell: As depicted in Figure 5a, two TA systems may be coupled to one another within a single cell by engineering plasmids in which the T expressed from plasmid 1 pairs with the A expressed from plasmid 2, and vice versa. Making defined variations on this scheme will enable us to make specific changes to the topology and/or parameters characterizing these MH networks. Variations to be constructed include: (a) Varying the copy number of each plasmid by changing its origin of replication; this would modulate network *states* by varying the stoichiometry of TA components. (b) Coupling copy number of plasmids 1 and 2 by utilizing the same origin of replication; because copy number is regulated by negative feedback operating at the origin of replication, placing both plasmids under the same origin of replication would couple their copy numbers in a new way and effectively *add a link or edge* to the network. (c) Expressing identical or distinct antibiotic resistance genes (Res) from each plasmid; by imbuing the two plasmids with distinct Res genes, selection with both drugs would constrain the observed distributions of plasmids 1 and 2 to those that confer resistance to both drugs (independent of TA interactions); by instead placing both plasmids under the same Res gene, the above constraint would be alleviated, such that TA dynamics would play a more dominant role in determining viable plasmid distributions, effectively *increasing the relative weight of edges* describing TA coupling.

Coupled TA networks between cells in a population: Although cells in a population always interact by competing for resources and generating waste products, an additional, explicit intercellular level of coupling may be added by engineering cells to contribute and respond to a common pool of a signaling molecule (Fig. 5b). Specifically, we will engineer cells to produce and respond to the well-characterized quorum sensing system from *Vibrio fischeri* used in many synthetic biology applications (81, 85). In this system, the enzyme LuxI catalyzes the synthesis of acyl-homoserine lactone (AHL), which diffuses into the intercellular milieu and into other cells. Intracellular AHL then binds to the LuxR protein, and the LuxR-AHL complex drives transcriptional activation of the *luxI* promoter. Thus, by making cells contribute and respond to the common pool of soluble factor AHL, this modification introduces a new hierarchical level of (intercellular) interactions. The strength of this coupling (i.e., the *weight of this edge*) can be controlled by making defined genetic modifications. For example, reducing the efficiency with which a given plasmid drives production of the enzyme LuxI would make such a cell more reliant upon the shared pool of AHL to induce genes regulated by LuxR (e.g., TA components, in the example depicted in Figure. 5b). Variations in network *topology* are also possible within this framework by making, for instance, one plasmid

independent of AHL regulation (the link representing the influence of the shared AHL pool would thus connect directly to the regulation of TA expression from one plasmid but not the other).

Using this family of experimental handles, we will generate libraries of plasmid "parts" that enable us to rapidly construct experimental scenarios representing network topologies with different levels of modularity and hierarchical organization, including the representative (but not exhaustive) scenarios depicted in Figure 5c. Notably, we have extensive experience at the Leonard Lab with engineering libraries of hundreds of synthetic biology constructs (3); a suitably broad variety of networks configurations will be constructed on a timescale of months using the basic parts described above.

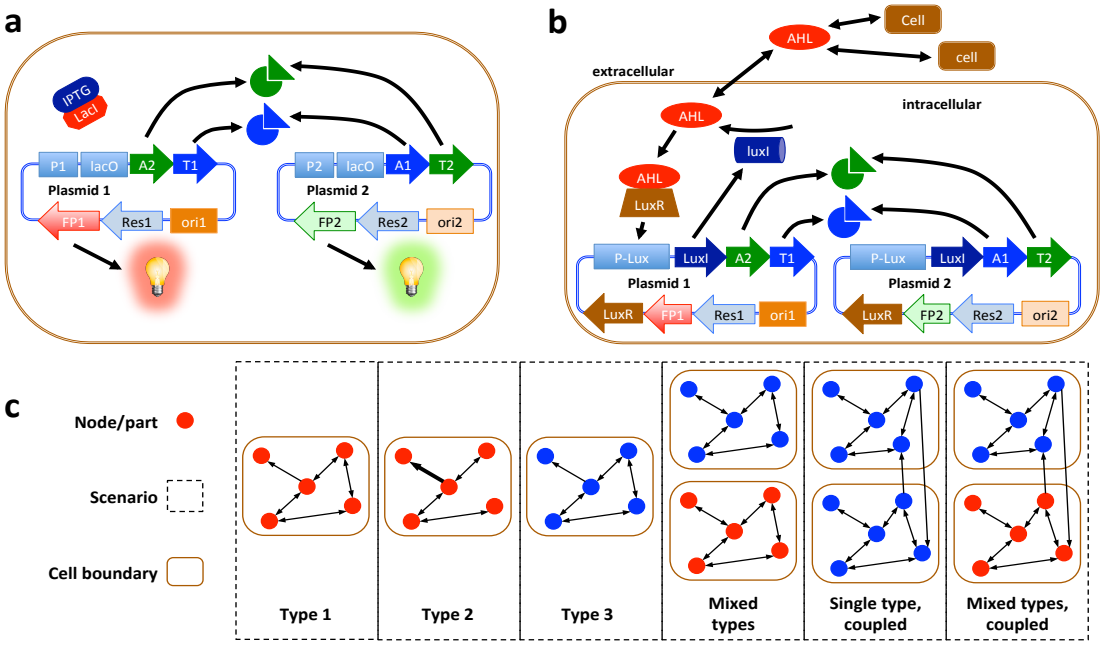

Figure 5. Experimental generation of model MH networks. (a,b) Biological implementation of networks in which (a) 2 different TA-encoding plasmids are coupled within a single cell and (b) multiple cells are coupled via secreted factors. (c) Abstract representation of variations and scenarios to be constructed and quantitatively characterized. Biological details: Res (antibiotic resistance); ori (origin of replication); LacI (IPTG drug-inducible switch to express TA components by removing LacI from lacO); AHL (soluble signal; see text); FP (fluorescent protein); P1,P2 (promoters).

Aim 1.2 Generate library of abstract MH networks and develop diagnoses for comparing them to experimental interaction networks. We will first focus our theoretical analysis on topology and study how steps 1, 2, 3, and 5 of our GA can be generalized by using different connectivity protocols (i.e. different N_m, $\mathbf{W}_=$, $\mathbf{W}_+$, S_-, $S_=$, and S_+) to produce a library of MH networks with different sizes, properties, and number of levels. We will apply various network analysis tools from the literature to these (multi-level) model networks and to our (comparatively small) engineered experimental networks to characterize them and develop common classification schemes that identify the hierarchical levels and degree of modularity. We will consider how these structures manifest themselves in the real world by studying their resiliency to noise and to alternative network descriptions such as grouping sets of nodes, switching between directed and undirected nodes, or having access to only partial connectivity information.

Aim 1.3 Characterize the fitness of experimental MH systems. To provide quantitative metrics for characterizing experimental MH system performance, we define two measures of "fitness": (1) maximum growth rate and (2) maximum biomass generated for a given media composition. In order to quantify these properties, microbial cultures comprising constructs from Aim 1.1 and representing scenarios such as those shown in Figure 5c will be inoculated in liquid growth medium, and the growth of these cultures over time will be monitored through optical density. Similar experiments will be conducted using rich and poor growth media to generate an initial template of fitness metrics for each MH network configuration.

Aim 1.4 Characterize the robustness of experimental MH systems. A collection of MH networks exhibiting high fitness in **Aim 1.2** will be further investigated to examine whether these topologies are robust to perturbations in network function (without changing network topology). For example,

translational efficiencies of T and A components may be individually perturbed in a predictable fashion (as described above) to explore highly related cases not generated in **Aim 1.1**. Robustness will defined as degree to which high fitness is maintained over a range of such perturbations.

Aim 1.5 Relating topology and dynamics in models and experiments. We will first study the connection between topology and dynamics in our model networks. We will implement the Boolean-based and ODE-based dynamics of node states described in **Section 4.2**. Boolean systems will allow us to systematically explore all possible dynamics supported by specific network structures. ODE systems will allow us to describe biological processes more realistically. The resulting dynamical properties will be included in the classification schemes developed in **Aim 1.2**.

We will then implement network-based and NSD-based representations of our experimental systems. Developing these models is non-trivial since it involves determining if the various types of indirect couplings present in real systems (e.g., competition for resources) should be explicitly represented or lumped into "black box" nodes. It will also require characterizing how the activity threshold used to establish links affects the resulting network topology. We will deduce and utilize a parsimony criterion to identify minimal network representations that properly model the experimental networks under a variety of conditions, based on the explicit fitness and robustness metrics quantified in **Aims 1.3-1.4**.

Aim 2. Explore the dynamics supported by MH structures

In this Aim we will systematically study the properties of the dynamical processes supported by the theoretical and experimental MH systems developed in **Aim 1**. We will identify the dynamical features that result from having MH structures by systematically studying processes on various fixed topologies with different degrees of MH organization. This research will also serve as a bridge between the network generation and characterization described in **Aim 1** and the full analysis of combined topological evolution and state dynamics proposed in **Aim 3**. Key questions are:

- What characterizes the dynamics of processes on MH systems and their response to external stimuli? Which theoretical structures best capture the dynamics of our experimental MH systems?
- Are the dynamics of network states different at different levels of an MH structure? Is there a time-scale separation that emerges directly from the MH interaction topology (as claimed in (114))?
- How do perturbations of the dynamics at one level of the hierarchy of an MH structure propagate within it and to other levels?
- What patterns are generated by NSD models and by experiments where cells on a substrate have intracellular interaction networks that can interact with each other via diffusion?

Aim 2.1 Experimental MH dynamics. The experimental tools constructed in **Aim 1** enable us to track the composition of the different nodes in a microbial network over time. In this subaim we will characterize these dynamics in detail. In particular, each plasmid-based construct will be engineered to drive expression of a distinct fluorescent protein (Fig. 5a) in order to track the relative ratio and content of such plasmids, including single cell-analyses subject to the limits of experimental noise (115). Moreover, this analysis will be bolstered by using population-averaged metrics such as qPCR to track the ratios of plasmids to one another and to genomic DNA, for which we have previously established a suitable set of methodologies (Fig. 2, (4, 5)). We will conduct growth experiments and use techniques described above to monitor dynamics at different levels of the hierarchical structure and under various conditions, including different network configurations (Fig. 5c), initial population compositions (when multiple cell types are used), and growth medium. Even simple microbial growth experiments can generate non-trivial and informative population dynamics resulting from interactions at the intercellular network level (116).

Aim 2.2 Theoretical MH dynamics. We will study in detail the dynamics of network states on the networks generated through the GA in **Aim 1**. We will implement ODE- and Boolean-based processes on them, as in **Aim 1.5**, and extract their dynamical properties. By considering a set of archetypical processes on the network, we will study how perturbations to stationary or periodic solutions propagate between modules and across the different levels of the network hierarchy. We will compare these results to those obtained for networks lacking MH structures and to experiments in **Aim 2.1**. We are particularly interested in exploring whether MH structures insulate modules from propagating perturbations or whether averaging effects within modules confers additional robustness to their resulting dynamics.

Aim 2.3 Diffusive and network interaction dynamics. We will study the dynamics of NSD models and compare them to experiments where cells interact via the exchange of soluble factors (i.e., AHL; see **Aim**

1.1). We expect NSD models to provide a rich platform for theoretical studies, since the spatial dynamics of a diffusively coupled array of networks can be as complex as that of reaction-diffusion systems or excitable media, potentially exhibiting pattern formation, multiple timescales, complex front or defect dynamics, and spatiotemporal chaos. We will develop numerical and analytical descriptions of these phenomena and generate predictions, such as pattern formation, that will be tested in corresponding experimental setups (e.g., growth of cell colonies on 2-dimensional agarose surfaces across which AHL can diffuse to link colonies). By providing experimental and theoretical insights into pattern formation and dynamics, this work could elucidate potential roles for MH structures in the evolution of morphogenesis.

Aim 3. Investigate the evolvability and adaptability of MH structures

Building upon the products of **Aims 1 and 2**, we will investigate how different network architectures impact a system's ability to (a) adapt and (b) evolve under changing conditions. In our theoretical framework, we can most directly interpret changes in network topology or in the rules that define the dynamics of network states as genotype changes, and changes in the network states as phenotype changes, both of which can be studied through numerical simulations. In our experimental framework, evolving the genotype would present an additional set of challenges, since it would involve implementing an evolutionary process that can induce and select mutations within a reasonable timescale. Our studies of adaptability can avoid dealing with these difficulties, while still investigating the effect of MH structures on fitness. Indeed, at the abstract level of our network descriptions, topological changes can correspond to either actual genotype variations that fundamentally change the biological processes or to changes in their operational dynamics also resulting in a different network representation. Therefore, having an evolvable system or an adaptable process may result in the same topological features.

In this Aim, we will study the evolvability and adaptability of MH systems. Our theoretical work will fully implement evolutionary algorithms and artificial-life simulations designed to study the emergence and stability of MH networks. Our experimental investigations will focus on the selection and adaptability of our synthetic biology systems engineered with different levels of modularity and hierarchical organization. Both efforts will be motivated by recent work showing that a variety of network systems evolve towards modular solutions when alternating fitness criteria are imposed (67-69). Specific question of this aim are:

- Do systems with higher modularity exhibit greater adaptability under changing environmental conditions or fitness criteria? What mechanisms can enhance modularity?
- Can we apply the principles resulting from the investigations above to evolve modular structures with two or more hierarchical levels spontaneously? Can synthetic biological systems be engineered that demonstrate the adaptability/evolvability of MH structures?
- How stable are MH structures from an evolutionary viewpoint? How do experimental and theoretical MH systems react to mutations or stresses imposed at different levels in their hierarchies?

Aim 3.1 Characterize the adaptability of MH networks. In this subaim, model experimental networks developed and characterized in **Aims 1 and 2** will be subjected to a novel experimental test: adaptability to changing environmental constraints. The selection of model networks used in this test will be guided by (a) their ability to demonstrate distinct levels of "fitness" and robustness (i.e. high, medium, low) when grown in a single growth medium in **Aim 1** and (b) theory-guided predictions of adaptability informed by the characterizations of the dynamics of MH systems in **Aim 2**. Experimental tests of adaptability will include the following perturbation scenario: microbial cultures will be sequentially exposed to media of different nutrient compositions (e.g., alternating rich/poor nutrient densities) using a microfluidic setup. We will use a modified commercial system (CellASIC) already established in the Leonard Lab, in which pneumatically-actuated flows enable precise control and dynamic modulation of the cellular microenvironment while continuously visualizing individual cells via fluorescent microscopy (i.e., to monitor expression of fluorescent proteins expressed from each TA plasmid). This will enable us to quantify individual cells' gene expression trajectories under constant or dynamic environments. We anticipate that shifting media composition will shift growth rate and thus the balance of various TA system components. It will also impact the degree to which various cells are coupled to one another as population density rises and falls. Additional experimental handles for perturbing environmental conditions will also be provided by our previous investigations into how oxygen limitations impact the dynamics of both cell growth and DNA content per cell, using a bioreactor platform in which oxygenation may be varied by altering the rate of sparging (4).

We will utilize the perturbations described above to investigate if structures (within the rich set of interaction networks produced in **Aim 1**) that have a higher degree of modularity achieve higher overall

fitness under shifting environmental conditions, and whether MH networks exhibiting high adaptability to one form of perturbation also exhibit high adaptability to different sorts of perturbations. We will model these experiments, using our theoretical systems to analyze the effect of the experimental handles and to explore this case of changing environmental conditions, which is complementary to prior numerical investigations of evolution under alternating fitness requirements and fixed environments (67-69).

Aim 3.2 Theoretical investigation of the evolution of MH structures. We will study the evolutionary dynamics of MH structures in network models by fully implementing our GA as an artificial life system. We will first generate an ensemble of networks with desired architectures and then iterate the GA to evolve under selective pressures imposed on the dynamics of network states. We will also study evolutionary dynamics on the NSD model. We will test models of evolution under alternating fitness criteria, mimicking the experiments proposed in **Aim 3.1** by introducing environmental perturbations as changes in the topology and dynamical rules. We will carry out various other numerical experiments, including:

Test if/how the results of the evolutionary algorithms depend on the network representation: Evolutionary dynamics can be implemented through the links or dynamical rules, and can be executed on various types of networks (directed or not, Boolean, ODE-based, etc.). A limit case of purely topological evolution can even be implemented using only NAND gates (the "universal" NOT-AND function with which any other logical function can be expressed). We will study how the evolved network structures depend on the specific description chosen, which representations are best suited for our experiments, and if there are unifying measures under which these disparate systems converge to the same results.

Carry out multiple artificial life experiments using Boolean and ODE-based dynamics: We will implement our Boolean or ODE-based dynamics on an ensemble of MH networks and make them compete under various fitness or environmental constraints. We will examine how the resulting features depend on over which nodes (and at which levels of the MH structure) we impose fitness or competition. We will consider the emergence of structures, studying under which conditions independent, symbiotic, or collaborative structures may emerge, and if these can be interpreted as a higher hierarchical level of modular organization (similar to an ecological network). The emergence of possible spatially segregated communities will be examined using the NSD model. If these are found, we will attempt to implement them experimentally. Finally, we will examine the mechanisms through which evolution under changing environmental or fitness requirements differ from evolution in under fixed conditions.

Characterize multi-level MH structures: We will also consider theoretical MH structures with multiple levels. We are interested in studying the evolution of substructures by implementing a version of the GA where, instead of evolving structures with a given number of levels, we continue adding levels while pruning nodes that do not contribute to the dynamics of nodes at the highest level. This constraint may impose conditions on the scaling properties of the overall emerging structure, in a way somewhat analogous to the Kolmogorov scaling-laws of fully developed turbulence in fluid dynamics (117-119), which can be deduced from the conservation of energy as it flows between eddies at different scales. We will explore which algorithms produce statistically stationary networks and study their properties.

6. Broader Impacts

The proposed research on how MH systems are structured, behave, and evolve could lead to breakthroughs in the beneficial utilization and manipulation of living systems, in the design of biologically inspired engineering solutions, and in our fundamental understanding of biological dynamics and self-organization. Our work could improve medical practice by helping identify critical nodes or structures in human or pathogen physiology that convey functional properties such as robustness. Enhancing or disrupting such motifs would generate novel therapeutic strategies. The project results could also inform ecological assessments and provide novel methods for understanding biodiversity, thus improving environmental stewardship. Understanding how evolution has shaped biological MH networks to achieve desirable properties could help us design engineered systems that also exhibit these features. Such systems could include engineered biological functions (built using the tools and methods of synthetic biology or other approaches) that enable the sustainable and cost-effective production of valuable chemicals, materials, or fuels. They could also include technological and social networks, where our increasingly interconnected world has a growing need for designs that are both robust and adaptable. Furthermore, our work will develop new approaches within nonlinear, statistical, and non-equilibrium physics that build upon the deep and extensive body of work in these areas to extend their applicability to MH systems and thus catalyze new contributions from these fields to the applications described above. Finally, our team will build upon our uniquely rich record of science outreach to conduct a series of

specific activities that leverage the results of the proposed research to engage, educate, and inform constituencies ranging from students to the public at large.

6.1 Self-organized modular hierarchical motion – a crowdsourcing experiment. In order to overcome the difficulty of communicating the somewhat abstract concept of MH systems, we will use a toy model that demonstrates intuitively the topological, dynamical and evolutionary properties of MH structures in an interactive simulation, also implemented as an online game and crowdsourcing experiment. This is a cornerstone activity in our plan to communicate the project perspective and results, and thus for enhancing its broader impact. It will uniquely integrate outreach and research by helping us (i) illustrate MH structures for a broad audience and build intuition through interactive simulations of mechanical MH systems, (ii) connect MH systems to engineering and robotic design, and (iii) implement crowdsourcing experiments to engage the public and compare designed MH structures to evolved ones.

The Active Elastic Membrane (AEM) model, recently introduced by Huepe et al, describes groups of self-propelled agents that are permanently connected by virtual springs to build larger structures similar to spring-mass models of elastic sheets (7, 11). Starting with random headings, agents self-organize into growing regions of coherent motion through a mechanism based on standard elasticity, eventually reaching collective rotation or translation (Fig. 6). We will extend our AEM model and simulation codes to MH structures by assembling a mechanical representation of Fig. 3, replacing nodes by self-propelled agents and links by springs. Components of embedded modules will self-organize into coherent motion and interact with other modules to also produce coherent motion in the embedding structures. This mimics MH dynamics in a system with interactions that induce ordered motion and whose only phenotypic trait is motion in physical space. Despite its simplicity, the MH-AEM model displays many of the features of MH systems.

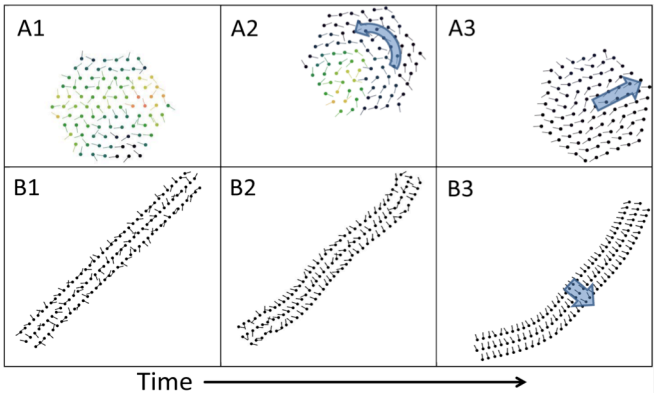

Figure 6. Simulations of active elastic membrane model. Collective motion and deformation spontaneously emerges from self-propelled components. Larger simulations of hierarchies of modular structures will demonstrate MH dynamics.

We will implement interactive simulations of MH-AEM structures composed of up to thousands of self-propelled agents that will illustrate visually MH dynamics and evolution and serve as a basis for our online game and crowdsourcing experiments. These simulations will place different AEM structures on a virtual arena to compete for randomly distributed resources or 'nutrients' in an artificial life setting. Since these AEMs self-organize into spontaneous motion of their whole and of their parts that is determined by their structure, their design will determine their ability to reach these nutrients more effectively than others. We will mutate these structures and impose selective pressures based on this ability to find nutrients or other motile capabilities and produce simulations where MH-AEM structures are visualized as they evolve based on the algorithms developed in **Aim 3**. These will help for outreach and to develop intuition.

We will also develop a crowdsourcing experiment based on these ideas, implementing an online game where participants are challenged to design MH-AEM structures that compete with other MH-AEM's (that were evolved or designed by other users) to achieve specific objectives. Note that many different virtual MH-AEM 'machines' can be engineered by adjusting the springs and self-propelled agents to create parts that self-organize to move, rotate, or deform in specific ways. The application will be designed to establish a library of modules created by users that can be combined to build more complex machines. MH design will thus naturally emerge, allowing a fascinating comparison between evolved and engineered solutions to the same problems. The AEM model has already been implemented as a Java application that will be set up in Amazon's Mechanical Turk or other crowdsourcing platform. This approach has the potential to generate widespread public interest (that we will encourage through press contacts), since it uniquely combines a mechanical analogy that illustrates the dynamical and evolutionary properties of MH systems for a broad audience with a game-based crowdsourcing experiment where participants can help our research by providing engineered solutions to evolutionary challenges.

6.2. Other outreach activities. We will maximize the impact of this work on applications deriving from diverse fields, by actively disseminating its results to audiences that span the biological sciences,

engineering, and physics. Dissemination will be enhanced through interactions with external collaborators (described in **Section 7**) and by the international nature of the project. We will target specialized physics and biology publications together with high-profile, broadly read journals such as Nature, Science, PNAS and PLoS. The postdoctoral researchers will present project results at the two largest international physics meetings, the March Meeting of the American Physical Society and the Spring Meeting of the German Physical Society, as described in the mentoring plan. We will also target presentations at the highly interdisciplinary International Conference on Network Science (NetSci). A project-specific website will be established to provide information for both professional scientists and interested non-professionals. The PI and co-PI's have a track record of press and media interactions and cultivate relationships with journalist at some of the most prestigious news outlets worldwide, which will facilitate the dissemination of project results to the broader public. In the past, this has led to news features in the New York Times, Wired Magazine, national television news, National Geographic and other high-profile media. Moreover, Dr. Huepe is also a professional musician involved in electronic music and in various art/science projects that reach a broad nonscientific audience.

We will also increase broader impact by enhancing STEM education. Prof. Leonard mentors Northwestern's iGEM team (120), which is developing ways to engage the public with synthetic biology, and we will integrate results from this project into that effort. He will also leverage this project to develop novel pedagogical materials for his interdisciplinary course on computational biology for undergraduate and graduate students. Prof. Brockmann will incorporate such material into lectures at the Institute for Biology (Humboldt-University) and in his position as Project Group leader at the Robert Koch Institute. Each of these efforts will generate novel pedagogical materials for integrating cutting-edge network biology research into the classroom, which will be shared in our public website for use by other instructors. Interfacing the project's research with teaching will also foster opportunities for undergraduates to get involved and pursue independent, related studies.

7. Project Management and Coordination

The project leaders have an established track record of collaboration and co-supervision of trainees. While all participants will be involved in the different aspects of the project, JL and the graduate student (GS) will lead its experimental aspects, DB and one postdoctoral research associate (PRA1) will lead the NSD approach, and CH and PRA2 will lead the MH network simulations and theory development. As described in detail in the **Postdoctoral Researcher Mentoring Plan**, the GS will join JL's lab at Northwestern University, PRA1 will be based in Berlin working primarily with DB, and PRA2 will be based at Northwestern University's Institute on Complex Systems (NICO) working primarily with CH. In addition, we have established a group of external (unfunded) collaborators, who will both provide consultation to ensure the success of the proposed work and catalyze the dissemination of its products to diverse disciplines. As described in attached letters (in Suppl. Docs.), expertise represented in this group includes adaptive-network based approaches (Prof. Thilo Gross), Boolean networks (Prof. Maximino Aldana), experimental synthetic biology (Prof. Calin Guet), cancer systems biology and medical applications (Prof. Sui Huang), and control of robotic systems and engineering applications (Dr. Ali E. Turgut).

Due to the international nature of this collaborative effort, effective project management and interactions will be maximized using the following strategy. Project coordination will involve frequent regular interactions among team members. Project-wide video conference meetings will be held every two weeks and include research updates from the PRAs and GS. We will also conduct biannual online milestone meetings that will include the external collaborators, where overall research progress, coordination, prioritization, collaborations and potential applications will be discussed, and upcoming challenges and opportunities will be identified and addressed. The tools and practices developed in previous experiences of successful long-distance and international collaborations by JL, CH, and DB (e.g. online collaborative resources) will be harnesses to ensure the successful integration of this project. Team members will also regularly interact in person: NICO facilities will host project-wide meetings to ensure strong interactions between the theoretical and experimental aspects of the project. CH is based in Chicago and is an established visiting researcher at NICO, while also spending 2 to 4 months every year at the Max Planck Institute in Germany (at no cost to the project). CH will thus be able to interact in person with DB's and JL's labs in Berlin and Chicago throughout the project, strengthening its cohesiveness. Finally, our proposed budget includes resources to enable the PRAs to visit annually the respective partner lab in Germany or Chicago, to facilitate research connections both within the project and between international partners more broadly.