

# Highly designable phenotypes and mutational buffers emerge from a systematic mapping between network topology and dynamic output

Yigal D. Nochomovitz\* and Hao Li†‡§

\*Graduate Group in Biophysics, †Department of Biochemistry and Biophysics, and ‡California Institute for Quantitative Biomedical Research, University of California, San Francisco, CA 94143

Edited by Michael Levitt, Stanford University School of Medicine, Stanford, CA, and approved January 14, 2006 (received for review August 18, 2005)

Deciphering the design principles for regulatory networks is fundamental to an understanding of biological systems. We have explored the mapping from the space of network topologies to the space of dynamical phenotypes for small networks. Using exhaustive enumeration of a simple model of three- and four-node networks, we demonstrate that certain dynamical phenotypes can be generated by an atypically broad spectrum of network topologies. Such dynamical outputs are highly designable, much like certain protein structures can be designed by an unusually broad spectrum of sequences. The network topologies that encode a highly designable dynamical phenotype possess two classes of connections: a fully conserved core of dedicated connections that encodes the stable dynamical phenotype and a partially conserved set of variable connections that controls the transient dynamical flow. By comparing the topologies and dynamics of the three- and four-node network ensembles, we observe a large number of instances of the phenomenon of “mutational buffering,” whereby addition of a fourth node suppresses phenotypic variation amongst a set of three-node networks.

designability | dynamical phenotype | enumeration | mutational buffering | regulatory network

Discerning the structure and function of cellular networks is essential to the development of a true understanding of biological systems. Experimental and theoretical studies are steadily advancing our knowledge of the wiring and input–output characteristics of a variety of natural and designed biological networks. These efforts have focused on characterizing the components and interactions for specific biological networks and their dynamical behaviors, by using standard genetic and biochemical approaches in conjunction with mathematical analysis of discovered circuits (1–7). Valuable insights into certain design features of biological networks have emerged through these efforts (8–13) and are used to guide the design of synthetic systems (14–16). The choice of which synthetic circuits to build is often a matter of careful hand-picking guided by experimental restrictions.

Reverse-engineering and modeling of specific experimental systems on a case-by-case basis is necessary and meritorious. However, the space of networks and associated dynamics is potentially very large, and parallel approaches that consider broad ensembles of networks may advance our understanding of general design principles in ways that the serial strategies may have difficulty revealing. Therefore, we have chosen to explore general design principles by using a global strategy. We expect that exploration of an entire ensemble of networks and associated dynamics will reveal statistical signatures connecting network architectures to categories of dynamical phenotypes. In particular, we analyze the relationship between the space of topology and the space of dynamics by employing an analogy to the protein “designability principle,” which states that compact protein structures that can be encoded by a wide array of sequences are preferentially selected in nature (17, 18).

Although the “designability principle” was originally proposed in the context of proteins (17–19), the concept is not limited to protein

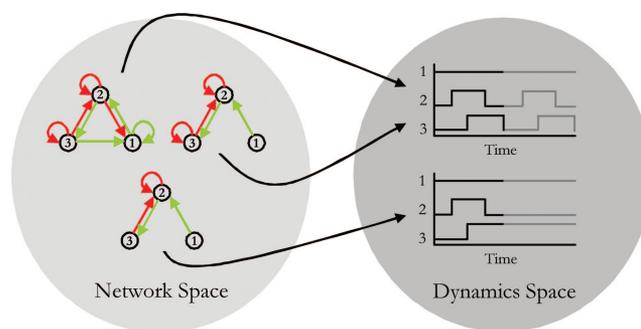


Fig. 1. Schematic mapping of designability for networks. The upper dynamical profile, a limit cycle that oscillates through the four states 100, 110, 111, and 101, possesses a designability of two in this mock example because the two upper left networks encode it. However, the lower dynamical profile arises from only one network, yielding a designability of one.

sequence structure relationships. In the most general setting, designability can be defined as the number of genotypes (i.e., sequences) that give rise to the same phenotype (i.e., structure). Here we explore such a generalization of the “designability principle” to biological networks, by defining the wiring diagram of node connections as the genotype that encodes the network’s dynamical output, i.e., its phenotype (Fig. 1). By using a simplified network model, we enumerated the entire space of network topologies and associated dynamical phenotypes. Certain dynamical phenotypes emerged as highly designable in that they can be encoded by a large number of network topologies. Comparison of networks of different size revealed many instances of the biological phenomenon of “mutational buffering,” whereby addition of a fourth node suppresses phenotypic variation among a set of three-node networks.

## Results

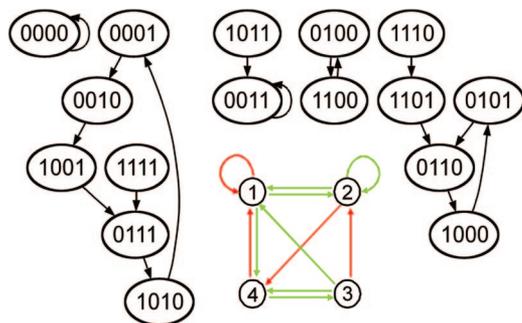
We adopt a simple Boolean model in search of general principles that govern the structure and behavior of small, three- and four-component networks. Network nodes represent discretized biological entities (mRNA, protein, etc.) that can be either on (1) or off (0), and the network connections represent regulatory relationships between the nodes. The state of a node evolves at discrete time steps according to the inputs from the other nodes and specific updating rules (see *Methods* for model description). We have sacrificed model complexity to permit exploration of an exhaustive ensemble of small networks, some of which may comprise modules of real biological circuits. We focus on three- and four-node networks because they are amenable to study by exhaustive enumeration in the context of our model and because reasonable evidence exists

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

§To whom correspondence should be addressed. E-mail: haoli@genome.ucsf.edu.

© 2006 by The National Academy of Sciences of the USA



**Fig. 2.** Phase space flow. The phase space for the four-node network depicted by using rule one. Green arrows, activating; red arrows, inhibitory. Binary string bits (left to right) represent the states of node one to four, respectively. We observe a 5-cycle with one transient (1111), a 3-cycle with two chained transients (1110, 1101), an isolated 2-cycle, a fixed point with one transient (1011), and a lone fixed point. Because the model contains no basal synthesis term, the state  $S^t = (0,0,0,0)$  is always a fixed point. There are a total of  $2^4 = 16$  states.

that biological circuits are comprised of small network modules linked together (20, 21).

In the network setting, one must define a metric to properly compare phase spaces. The spectrum of limit cycles and/or fixed points within phase space characterizes a network's stable behavior and thus defines its primary phenotype. Therefore, in our designability analysis we focus our attention on the presence of individual limit cycles and fixed points. We address the transient dynamics, which represent a compelling but secondary phenotype, later in *Results*. We define the “designability” of a dynamical phenotype, either a limit cycle passing through a specific series of states or a particular fixed point, as the number of topologies that contain that phenotype within the topology's dynamical phase space. For instance, in Fig. 2, the depicted network encodes a limit cycle containing five states ( $0001 \rightarrow 0010 \rightarrow 1001 \rightarrow 0111 \rightarrow 1010$ ); thus, this particular network contributes one to the designability of the 5-cycle. Another closely related topology (not shown) encodes the same 5-cycle; hence, we would increment this 5-cycle's designability to two. We continue adding one to the designability of this 5-cycle for each network in the four-node ensemble that generates the 5-cycle and the total counts after we have enumerated all possible networks gives the designability of the 5-cycle. For 3-cycles and higher, we treat opposing temporal orientations of limit cycles as distinct phenotypes.

**Overview of Dynamical Phenotypes.** Table 1 summarizes the scope of dynamical phenotypes that our exhaustive network enumeration captures. Here we report only the phenotypic coverage for the  $n = 3$  and  $n = 4$  ensembles when measured against the theoretical pool of possible  $k$ -cycles. We observe that for two different updating rules corresponding to different decay processes (see *Methods* for rule descriptions), as the cycle complexity (measured by increasing  $k$ ) rises, the ratio of  $k$ -cycles that can be designed by at least one topology to the number of theoretically possible  $k$ -cycles falls sharply (see also Fig. 3A). This observation draws our attention to the intermediate phenotypes that are neither excessively rare nor excessively simple. Such phenotypes exhibit the greatest phenotypic variation ( $k = 4, 5, 6,$  and  $7$  for  $n = 4$ ; and  $k = 2, 3,$  and  $4$  for  $n = 3$ ). We find that under rule two a smaller fraction of dynamical phenotypes are designable by at least one topology. The last two columns indicate that for  $n = 3$  and  $n = 4$  up to 5-cycles, the phenotypic coverage for rule two is a complete subset of the phenotypic coverage corresponding to rule one. However, for 6-cycles and higher, the increased complexity of the  $n = 4$  ensemble yields a small fraction of phenotypes that are unique to rule two.

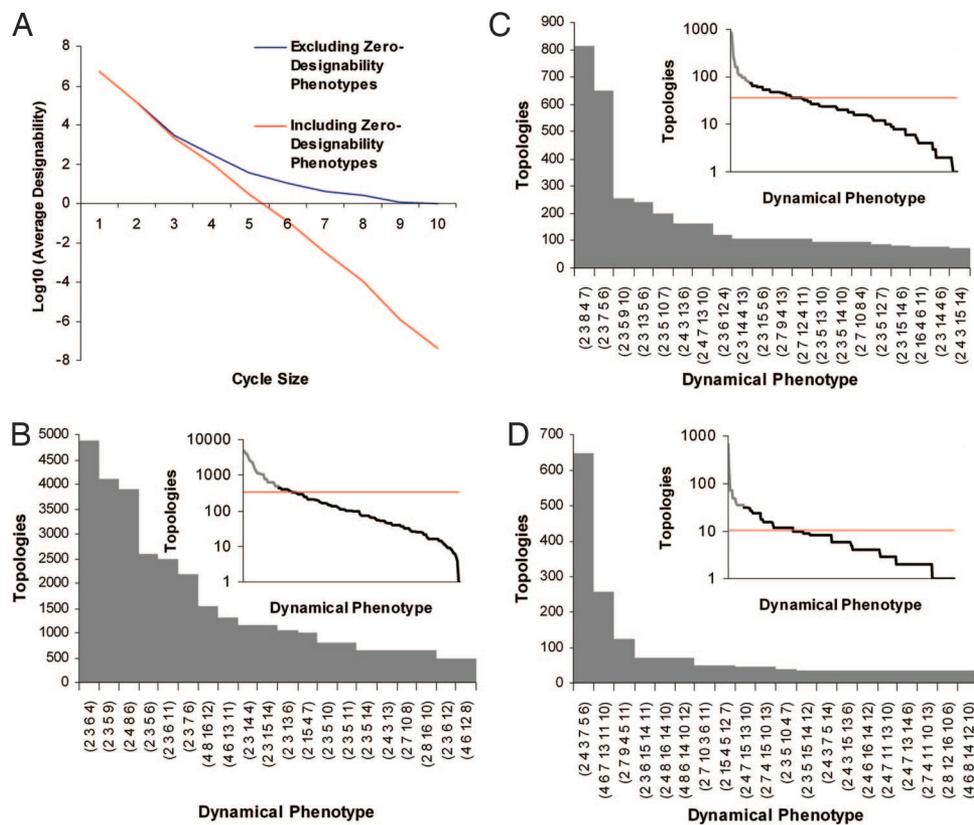
Whereas similar trends between the phenotypic coverages suggest a reasonable correlation between the two updating rules for both  $n = 3$  and  $n = 4$ , it is an appreciable correlation of designabilities that confirms that the two rules are statistically equivalent. We performed a correlation analysis by registering the designability scores for each phenotype reported in Table 1 (see Fig. 8, which is published as supporting information on the PNAS web site). The correlation is highly significant as judged by linear regression ( $R^2 = 0.836$ ). Although the microscopic rank-ordering of designabilities differs somewhat under each rule (data not shown), we conclude that for the phenotypes which arise under both rules, the strong, intermediate, and weakly designable phenotypes emerge consistently.

**Certain Dynamical Phenotypes Are Highly Designable.** We now describe results from a more detailed analysis. Given the macroscopic equivalence of rules and the peak in phenotypic variation for the intermediate-sized limit cycles, we focus on rule one for the remainder of this paper and intermediate  $k$ -cycles for four-node networks. Qualitatively similar results were obtained for three-node networks (data not shown). Fig. 3A plots average designability against cycle size, indicating that designing larger cycles is exponentially hard (Fig. 3A, red line). When phenotypes that cannot be designed at all by our model are excluded from the average (Fig. 3A,

**Table 1. Coarse-grained phenotype analysis**

$k$	Theoretical pool of possible $k$ -cycles		$k$ -cycles that occur under rule 1		$k$ -cycles that occur under rule 2		$k$ -cycles that overlap rules 1 and 2	
	$n = 4$	$n = 3$	$n = 4$	$n = 3$	$n = 4$	$n = 3$	$n = 4$	$n = 3$
10	1,089,728,640	$\emptyset$	48	$\emptyset$	72	$\emptyset$	24	$\emptyset$
9	201,801,600	$\emptyset$	192	$\emptyset$	0	$\emptyset$	0	$\emptyset$
8	32,432,400	$\emptyset$	1,476	$\emptyset$	486	$\emptyset$	390	$\emptyset$
7	4,633,200	720	3,600	0	1,176	0	983	0
6	600,600	840	6,776	2	2,020	2	1,900	2
5	72,072	504	5,688	12	1,776	6	1,776	6
4	8,190	210	2,730	24	1,158	18	1,158	18
3	910	70	668	34	404	10	404	10
2	105	21	105	21	77	12	77	12
1	16	8	16	8	16	8	16	8

Theoretical counts of the total number of possible  $k$ -cycles compared with  $k$ -cycles that can be designed by at least one topology for rule one, rule two, and both rules. No  $k$ -cycles larger than 10 occur in the  $n = 4$  ensemble. The symbol  $\emptyset$  indicates that  $k$ -cycles 8, 9, and 10 are *a priori* mathematically impossible for  $n = 3$  because there are only  $2^3$  states and the fixed point  $S^t = (0,0,0,0)$  cannot participate in a limit cycle. Theoretical counts were computed according to the formula  $\binom{2^n - 1}{k} (k - 1)!$ .



**Fig. 3.** Certain dynamical phenotypes are highly designable. (A) Average designability vs. cycle size: Designing larger cycles is exponentially hard (red line). When zero-designability phenotypes are excluded (blue line), average designability follows subexponential behavior, tending to one. (B–D) Designability spectra for 4-node networks: B, 4-cycles; C, 5-cycles; D, 6-cycles. The x axis represents dynamical phenotypes ordered by designability. The gray portions of the full spectra (insets) correspond to the 20 most designable  $k$ -cycles, shown in the main plot. Horizontal red lines in the insets record the mean designability for each  $k$ -cycle. Key to dynamical states: 2 = 0001, 3 = 0010, 4 = 0011, 5 = 0100, 6 = 0101, 7 = 0110, 8 = 0111, 9 = 1000, 10 = 1001, 11 = 1010, 12 = 1011, 13 = 1100, 14 = 1101, 15 = 1110, 16 = 1111. For example, (2 3 6 4) represents 0001 → 0010 → 0101 → 0011 → 0001.

blue line), the average designability is bounded from below by one and the behavior becomes subexponential.

Fig. 3 B–D depict designability spectra for 4-, 5-, and 6-cycles, respectively. Gauging from Fig. 3A, the average designabilities decrease rapidly as the length of the  $k$ -cycle increases. Therefore, it is necessary to distinguish bulk designability scalings that arise from mere cycle complexity from designabilities that reflect the underlying structure of the dynamical phenotypes. We therefore segregated the cycle sizes into independent plots to compare limit cycles of equal complexity.

Each inset in Fig. 3 displays the full designability spectrum in a semilog graph in which the light-gray portion of the spectrum, corresponding to the 20 most designable  $k$ -cycles, is expanded in the main plot. Such “highly designable” phenotypes substantially exceeded the mean designabilities (Fig. 3 Insets, horizontal red lines) of 322.4, 35.6, and 10.7 for the 4-, 5-, and 6-cycles, respectively. In contrast, for each  $k$ -cycle, we observe a long tail indicated by the black portion of the inset spectrum that is comprised of weakly designable phenotypes.

The designability spectra are highly nonrandom. To illustrate, among the 5-cycles, the probability of observing the top-ranking designability of 810, assuming that designabilities are governed by a Poisson process with mean 35.6, is  $1.32 \times 10^{-763}$  (see Fig. 9, which is published as supporting information on the PNAS web site). The emergence of strongly preferred dynamical phenotypes against a wide background of less successful ones suggests that a “designability principle” may be operative for networks.

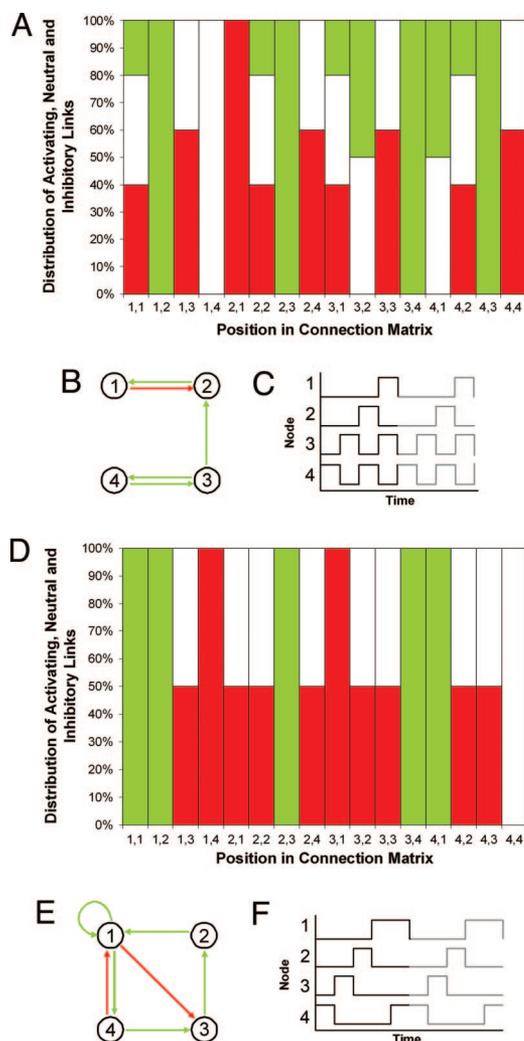
**Topological Variability: Two Case Studies.** What dynamical features lead to the success of these highly designable dynamical phenotypes, and what are the characteristics of the underlying topologies of the corresponding networks? Although we cannot give insights into every phenotype and its associated topologies, we present as case studies two highly designable examples.

We examined the variability of each topological connection for

a set of networks that design a given phenotype. Specifically, using the connectivity matrix to represent the topology, we counted the number of activating, inhibitory, and neutral links for each matrix position across the set. We discovered two classes of connections: fully conserved and variable. The fully conserved connections correspond to a subnetwork module that is necessary to encode the designable, stable phenotype. The variable connections control the transient dynamical flow. The partitioning of connections into a fully conserved core and a variable set may suggest a framework for understanding the mapping from topological space to dynamical space.

Fig. 4 A and D each depict an example of topological variability. Vertical bars of a single color correspond to fully conserved positions of the connectivity matrix. Fig. 4A, for instance, depicts the topological variability for a highly designable 4-cycle phenotype whose dynamics are illustrated in Fig. 4C. There are 6 fully conserved positions (four activating, one inhibitory, and one neutral) and 10 variable positions that display biases for particular link types, variations that collectively give rise to 2,500 topologies. We observe moderate dependencies among the choices at the 10 variable positions: if the selections at each variable position are independent, one should observe  $2^6 \times 3^4$  or 5,184 networks, yet only 2,500 occur.

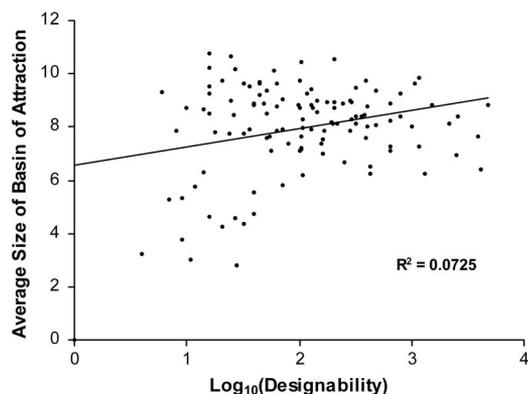
The fully conserved activating links generate a cascading pattern of node activations: node four turns on node three, which in turn switches on node two, which finally turns on node one and the cycle repeats (Fig. 4C, light-gray portion). The cascade is reinitiated by node four at every other time step because there is a bidirectional activation between nodes three and four. Furthermore, the class of networks presented in this example possesses one specialized connection: a fully conserved inhibitory link from node one to node two. Therefore, the attempt to generate a full cascade originating from node four falls silent at node two half of the time, leading to off spans for nodes one and two that are 3-fold longer than those for nodes three and four.



**Fig. 4.** Topology variability analysis. (A) Highly designable 4-cycle (2 3 6 11). (D) Highly designable 5-cycle (2 3 5 9 10). (B and E) Network diagrams displaying only the fully conserved positions of the connectivity matrix. Green, activating; red, inhibitory; white, neutral. (C and F) Dynamics for two oscillations (second oscillation in gray) of each limit cycle.

Fig. 4 D–F depicts another highly designable dynamical phenotype, a 5-cycle that generates an uninterrupted cascade. Here the fully conserved core uses four specialized links to produce a different modulation of a cascading pattern of node activations. The autoactivating link on node one, the activation from node one to node four, and the two inhibitory links act in concert to keep nodes one and four on twice as long as nodes two and three. In this circumstance, there are no dependencies at the variable positions because the designability is precisely 256. Weakly designable dynamical phenotypes necessarily possess a greater fraction of fully conserved connections and contain increasingly idiosyncratic features that belie straightforward categorization (see Fig. 10, which is published as supporting information on the PNAS web site).

**Designability Is Independent of Robustness.** Each stable phenotype possesses a basin of attraction: the collection of all transient states that flows toward the stable phenotype. The size of the basin of attraction combined with the size of the  $k$ -cycle measures a stable phenotype's robustness to environmental perturbations. The correlate of robustness for proteins is thermodynamic stability. For protein designability on lattices, one can frame a geometric argument for the positive correlation between highly designable folds



**Fig. 5.** Designability and the average size of the basin of attraction are uncorrelated. The analysis was performed for the 2,730 4-cycles arising from four-node networks under rule one.

and thermodynamic stability (18). We were curious to know whether an analogous correlation holds between robustness and designability in the network setting.

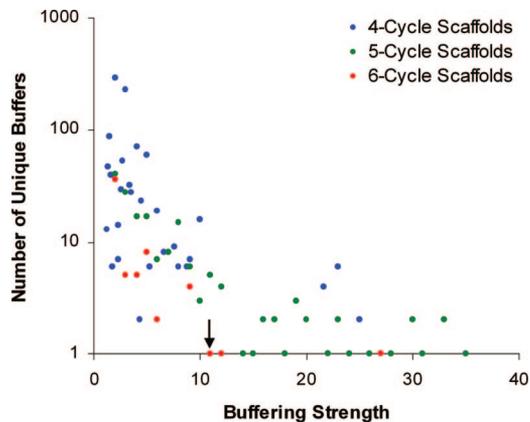
We performed a correlation analysis between a phenotype's designability and the average size of that phenotype's basin of attraction (Fig. 5) and discovered no significant correlation. The absence of a correlation holds true for all  $k$ -cycles under either updating rule for both the  $n = 3$  and  $n = 4$  ensembles (data not shown). This result may reveal an interesting deviation from the designability conclusions for proteins.

**Mutational Buffering.** We conducted a systematic exploration of the interplay between the mappings from topological to dynamical space for the three- and four-node network ensembles. Our motivations for this analysis are drawn from the genetic concept of mutational buffering. Mutational buffers assist in preserving a common phenotype by suppressing the phenotypic effect of genetic variations. We therefore searched for circumstances where the incorporation of a fourth node bearing specific connections to a set of genetically distinct three-node networks would trigger the three-node networks to acquire the same dynamical phenotype, despite possessing differing dynamical phenotypes as isolated three-node systems.

To investigate this idea, we systematically searched the four-node  $k$ -cycle and  $2k$ -cycle dynamics for matches (on three of the four nodes) to  $k$ -cycles generated by three-node networks, holding  $k$  fixed for any particular search ( $2k$ -cycles are included because they could possess embedded  $k$ -cycles orbiting twice on three of the four nodes). We then imposed the criterion that the fourth node demonstrate a fixed set of connections, i.e., a scaffold, with respect to the original three-node network. We also fixed the identity of the fourth node's self-coupling (only activating, only inhibitory, or only neutral) and demanded that the fourth node's dynamics display activity, i.e., not trivially off. Therefore, for any particular scaffold, the fourth node exhibits a specific and active functional role. For each scaffold identified, we divided the number of four-node networks that design the  $k$ -cycle using the specified scaffold by the number of three-node networks in isolation that also design the  $k$ -cycle. The quotient measures designability amplification, or buffering strength (Fig. 6,  $x$  axis).

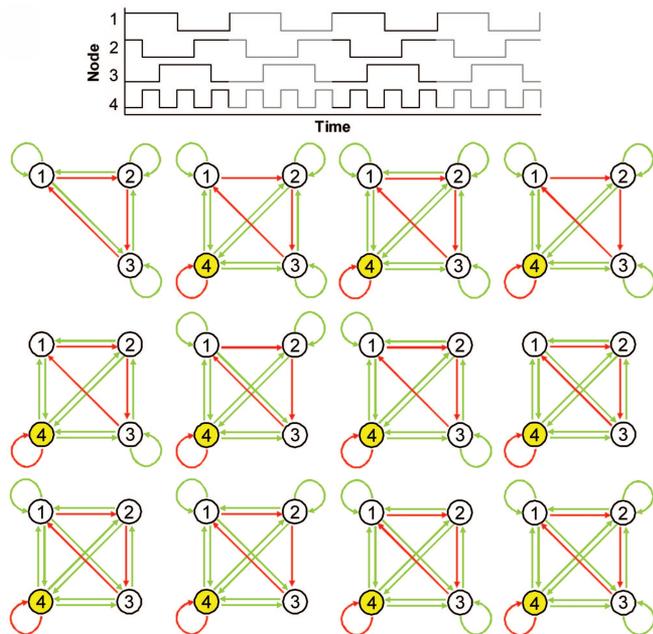
In Fig. 6 we report the results of the systematic search for 4-, 5-, and 6-cycle scaffolds. Amongst these  $k$ -cycles, we observe 1,356 instances of the scaffolding phenomenon with an average buffering strength of 3.63. Most scaffolds confer buffering strengths in the lower ranges. However, we also discovered instances of high buffering strengths, for example an amplification of 35 for one 5-cycle.

To portray a simple example, we examine in detail the unique



**Fig. 6.** Systematic search for mutational buffers. A scaffold refers to a fixed manner by which a fourth node interacts with a three-node subnetwork. Buffering strength reports the ratio of designability of a dynamical phenotype observed on three-node subnetworks attached via scaffold to a fourth node, to that phenotype's designability on isolated three-node networks. The black arrow points to the example portrayed in Fig. 7. Dynamics were generated by rule one.

6-cycle scaffold that possesses a buffering strength of 11 (Fig. 6, black arrow). The network at the upper left of Fig. 7 is the only member of the three-node ensemble that can design the shifted “three-on/three-off” pattern of node activities shown in the dynamical profile for nodes one, two, and three appearing at the top of Fig. 7. However, by incorporating a fourth node (Fig. 7, yellow node) that interacts with each original node via bidirectional activations, 11 unique four-node networks arise that support the shifted “three-on/three-off” dynamics of the isolated three-node network.



**Fig. 7.** A mutational buffer in action. The three-node network at the upper left encodes the 6-cycle phenotype shown above (for nodes one, two, and three) that possesses a designability of one. The fourth yellow node interacts with the original three-node network in a fixed manner via three sets of bidirectional activations. By virtue of these interactions, 11 four-node networks can now confer the 6-cycle phenotype to nodes one, two, and three, resulting in a buffering strength of 11 (Fig. 6, black arrow).

Ten of these subnetworks, in isolation, would be unable to support the dynamical profile of the original three-node network [the 11th subnetwork (lower right in Fig. 7) is identical to the original three-node network]. The incorporation of a fourth node buffers the subnetworks' internal variations, restoring their capacity to produce the dynamics of the isolated three-node network. The “buffering” node requires a unique dynamical signal, oscillating at three times the frequency of the nodes in the subnetwork (Fig. 7, dynamics profile for node 4) to compensate in a consistent manner for the alternate wirings of the subnetworks. The buffer elegantly restores bridging communication between nodes in the three-node subnetworks whose direct activating links have been mutated away when compared with the original three-node network (Fig. 7, upper left).

## Discussion

We have conducted an exhaustive computational analysis of the mapping from topological to dynamical space for a simplified network model that abstractly mimics elements of biological regulatory networks. Our investigation complements theoretical studies that eschew direct enumeration in favor of providing certain analytical insights into the connections between network topology and dynamics (22–24).

Our principal conclusion is that a small fraction of dynamical phenotypes possesses atypically high designabilities. This finding conceptually parallels the protein case, where a small subset of folds in simple two- and three-dimensional enumerative models possesses unusually high designabilities (17–19). Highly designable dynamical phenotypes arise via the presence of a core of fully conserved network links. The design of the core is effective because the core is relatively small but can tolerate wide variations in the link identities at the residual nonconserved positions while keeping the encoded stable phenotype intact. The presence of a conserved network core corresponds conceptually to protein designability, where a subset of essential residues defines the fold and variations at the nonessential residues absorb a large spectrum of sequences (17–19). Despite these observations, we still seek a general theoretical understanding of the topological features of the conserved network cores that enable them to yield highly designable dynamical phenotypes.

The variable links determine the pathways of the transient states to the stable phenotypes and modulate the stable phenotypes' robustness to environmental perturbations. Therein lies a superficial correspondence to proteins, in that the variable residues specialize details of the folding pathway and/or confer differential stabilities to equivalent folds. However, unlike proteins, where a correlation exists between stability and designability, no such correlation was observed between the stability of a dynamical phenotype (as measured by the size of the basin of attraction) and its designability in our network model. In contrast, it seems that the structure of the dynamical phenotype itself, rather than the phenotype's capacity to accommodate a large collection of transient states, contributes principally to high designability.

We have also shown that a Boolean model can provide mechanistic insight into how dynamical buffering may arise. Our systematic comparison of the three- and four-node ensembles' topologies and dynamics lends theoretical support to the possibility that similar dynamical phenotypes occurring in gene regulatory networks of different strains may be preserved via mutational buffers. We discovered that dynamical buffering could exhibit complex, nonintuitive features, demonstrating that the adoption of enumeration methodologies to probe design principles of biological networks is useful in revealing designs that may not be found by intuition. Other computational investigations into the buffering phenomenon (25, 26) have yielded related conclusions but provided no detailed mechanistic insight into the action of identified buffers.

Mutational buffering is common in biology (27, 28). The prospect that buffering could provide a measure of phenotypic robustness to

internal genetic variation amongst different strains through intricate dynamical relationships could expand the view of mutational buffering as a phenomenon traditionally attributed to static genetic dominance or redundancy. A buffer acting to suppress phenotypic variation through specific dynamical relationships to other genes is an interesting prediction that could be tested experimentally. Although we could only portray one example in *Results*, we have elucidated >1,000 buffering scaffolds for small, experimentally accessible four-node systems. For each scaffold, we have produced a complete map of the spectrum of genetic variation that the scaffold suppresses, the phenotype that the scaffold preserves, and the dynamical behavior of the scaffolded buffer itself. These examples can be used to guide the selection and design of specific networks to test the idea of mutational buffering via intricate dynamic regulation (14).

There are advantages and limitations to Boolean modeling. The simplicity of a Boolean model allows us to enumerate the full space of topology and dynamics, which leads to observation of phenomena (such as the existence of highly designable dynamical phenotypes and mutational buffers) that are impossible to reveal by case studies. In certain instances, Boolean modeling accurately recapitulates the features of dynamical processes observed by using differential equations that represent plausible biological regulatory mechanisms. For example, we have demonstrated (see Fig. 11, which is published as supporting information on the PNAS web site) that the networks portrayed in Fig. 7 possess Boolean dynamics with the same phase-shifted oscillations and frequency ratios that arise when the dynamics are controlled by differential equations. In other circumstances, the Boolean approach suffers from nonphysical artifacts that are detected by continuous models (29).

Our designability study for networks will guide experimentalists and theorists toward consideration of statistically preferred dynamical outputs in addition to analysis of statistically overrepresented network motifs (20). Although the results we presented here are limited to toy models, we anticipate that the concepts of designability of dynamical phenotypes and mutational buffering will become valuable in understanding the design of real biological networks.

## Methods

We define a network topology according to an  $n \times n$  connectivity matrix,  $C$ . Each matrix element  $C_{ij}$  represents the connection from node  $j$  to node  $i$  and is chosen from the set  $\{-1, 0, 1\}$ , corresponding to inhibition, neutral, and activation. The model hence permits bidirectional links, autoactivations, and autoinhibitions. For a network of  $n$  nodes, the model yields  $3^n$  networks upon exhaustive enumeration of topologies ( $n = 3$ ,  $3^3 = 19,683$ ;  $n = 4$ ,  $3^4 = 43,046,721$ ). We implement the analytic power of the Pólya enumeration theorem to generate an exhaustive set of nonisomorphic

graphs (see *Supporting Text*, which is published as supporting information on the PNAS web site). After pruning isomorphic graphs, we are reduced to 3,411 ( $n = 3$ ) and 1,809,459 ( $n = 4$ ) unique topologies.

We use a relatively simple procedure to explore the dynamical behaviors of a large ensemble of networks. For each discrete time point  $t$ , the state of each node  $i$ , denoted  $S_i^t$ , can be either 1 (on) or 0 (off). The input to each node is given by  $\sum_{j=1,n} C_{ij} S_j^t$ . The assignment of the next state  $S_i^{t+1}$  is then determined by the net input:  $S_i^{t+1}$  is mapped to one for net activation and zero for net inhibition. The nodes are synchronously updated at each time step. In the cases where the net input to a node is zero, we consider two updating rules,  $\mathfrak{S}_1$  and  $\mathfrak{S}_2$ . Under rule one ( $\mathfrak{S}_1$ ), the state of a node presented with a zero net input at time  $t$  decays to zero or remains zero at time  $t + 1$ . Under rule two ( $\mathfrak{S}_2$ ), the state of a node presented with a zero net input at time  $t$  retains its previous state. The two rules are summarized as follows:

$$\mathfrak{S}_1: S_i^{t+1} = \begin{cases} 1 & \sum_{j=1,n} C_{ij} S_j^t > 0 \\ 0 & \sum_{j=1,n} C_{ij} S_j^t \leq 0 \end{cases} \quad [1]$$

$$\mathfrak{S}_2: S_i^{t+1} = \begin{cases} 1 & \sum_{j=1,n} C_{ij} S_j^t > 0 \\ 0 & \sum_{j=1,n} C_{ij} S_j^t < 0 \\ S_i^t & \sum_{j=1,n} C_{ij} S_j^t = 0 \end{cases} \quad [2]$$

The physical interpretation of  $\mathfrak{S}_1$  is that the time scales of the decay process and the activating or inhibitory processes are of equal order. The physical interpretation of  $\mathfrak{S}_2$  is that the decay process occurs on a much slower time scale than the activating or inhibitory processes.

To calculate a network's phase space, we compute the successor state for each of the  $2^n$  initial conditions and employ a graph-searching algorithm to deduce the cycle decomposition and transient flow (30). The space of dynamical outputs scales as  $(2^n)^{(2^n-1)}$ , which significantly outpaces the topology count of  $3^n$ . For  $n = 3$  and  $n = 4$  there are, respectively, 2,097,152 and  $1.15 \times 10^{18}$  theoretically possible dynamical phase spaces.

We thank Chen-Shan Chin, Jeffrey Chuang, Hana El-Samad, Surya Ganguli, Morten Kloster, and Chao Tang for valuable comments; and Alisa Nochomovitz for editing the manuscript. This work was supported by National Institutes of Health Training Grant T32 GM08284 (predoctoral support for Y.D.N.), National Institutes of Health Grant GM70808, the National Natural Science Foundation of China (No. 100228510), and a David and Lucile Packard Fellowship (to H.L.).

- McAdams, H. H. & Shapiro, L. (1995) *Science* **269**, 650–656.
- Jaspersen, S. L., Charles, J. F., Tinker-Kulberg, R. L. & Morgan, D. O. (1998) *Mol. Biol. Cell* **9**, 2803–2817.
- Li, F., Long, T., Lu, Y., Ouyang, Q. & Tang, C. (2004) *Proc. Natl. Acad. Sci. USA* **101**, 4781–4786.
- Ueda, H. R., Hayashi, S., Chen, W., Sano, M., Machida, M., Shigeyoshi, Y., Iino, M. & Hashimoto, S. (2005) *Nat. Genet.* **37**, 187–192.
- Alon, U., Surette, M. G., Barkai, N. & Leibler, S. (1999) *Nature* **397**, 168–171.
- Zhu, X. M., Yin, L., Hood, L. & Ao, P. (2004) *J. Bioinform. Comput. Biol.* **2**, 785–817.
- Bagowski, C. P., Besser, J., Frey, C. R. & Ferrell, J. E., Jr. (2003) *Curr. Biol.* **13**, 315–320.
- Morgan, D. O. (1997) *Annu. Rev. Cell Dev. Biol.* **13**, 261–291.
- Kobiler, O., Rokney, A., Friedman, N., Court, D. L., Stavans, J. & Oppenheim, A. B. (2005) *Proc. Natl. Acad. Sci. USA* **102**, 4470–4475.
- Ferrell, J. E., Jr. (2002) *Curr. Opin. Cell Biol.* **14**, 140–148.
- Berg, H. C. (1988) *Cold Spring Harbor Symp. Quant. Biol.* **53**, 1–9.
- Barkai, N. & Leibler, S. (1997) *Nature* **387**, 913–917.
- Kato, A., Latifi, T. & Groisman, E. A. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 4706–4711.
- Yokobayashi, Y., Weiss, R. & Arnold, F. H. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 16587–16591.
- Elowitz, M. B. & Leibler, S. (2000) *Nature* **403**, 335–338.
- Guet, C. C., Elowitz, M. B., Hsing, W. & Leibler, S. (2002) *Science* **296**, 1466–1470.
- Li, H., Helling, R., Tang, C. & Wingreen, N. S. (1996) *Science* **273**, 666–669.
- Li, H., Tang, C. & Wingreen, N. S. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 4987–4990.
- Wingreen, N. S., Li, H. & Tang, C. (2004) *Polymer* **45**, 699–705.
- Thieffry, D. & Sánchez, L. (2004) in *Modularity in Development and Evolution*, eds Schlosser, G. & Wagner, G. P. (Univ. of Chicago Press, Chicago), pp. 222–243.
- Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. & Alon, U. (2002) *Science* **298**, 824–827.
- Klemm, K. & Bornholdt, S. (2005) *Phys. Rev. E* **72**, 055101(R).
- Klemm, K. & Bornholdt, S. (2005) *Proc. Natl. Acad. Sci. USA* **102**, 18414–18419.
- Remy, E., Mosse, B., Chaouiya, C. & Thieffry, D. (2003) *Bioinformatics* **19**, Suppl. 2, ii172–ii178.
- Bergman, A. & Siegel, M. L. (2003) *Nature* **424**, 549–552.
- Siegel, M. L. & Bergman, A. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 10528–10532.
- Queitsch, C., Sangster, T. A. & Lindquist, S. (2002) *Nature* **417**, 618–624.
- Hartman, J. L., IV, Garvik, B. & Hartwell, L. (2001) *Science* **291**, 1001–1004.
- Glass, L. & Kauffman, S. A. (1973) *J. Theor. Biol.* **39**, 103–129.
- Tarjan, R. E. (1972) *SIAM J. Comput.* **1**, 146–160.