

## **Intracellular Regulatory Networks are close to Monotone Systems**

Avi Ma'ayan.<sup>\*</sup>, Ravi Iyengar<sup>\*</sup> and Eduardo Sontag<sup>†</sup>

<sup>\*</sup> Department of Pharmacology and Biological Chemistry, Mount Sinai School of Medicine,  
1 Gustave Levy Place, New York, NY 10029

<sup>†</sup> Department of Mathematics, Rutgers, The State University of New Jersey, Hill Center,  
110 Frelinghuysen Road, Piscataway, NJ 08854-8019

Corresponding Author:

Ravi Iyengar, Ph.D.

Department of Pharmacology and Biological Chemistry

Mount Sinai School of Medicine

1 Gustave Levy Place, Box 1215

New York, NY 10029, USA

E-mail: ravi.iyengar@mssm.edu

Tel: (212) 659-1700

Fax: (212) 831-0114

## **Abstract**

Several meso-scale biological intracellular regulatory networks that have specified directionality of interactions have been recently assembled from experimental literature. Directed networks where links are characterized as positive or negative can be converted to systems of differential equations and analyzed as dynamical systems. Such analyses have shown that networks containing only sign-consistent loops, such as positive feed-forward and feedback loops function as monotone systems that display well-ordered behavior. Perturbations to monotone systems have unambiguous global effects and a predictability characteristic that confers advantages for robustness and adaptability. We find that three intracellular regulatory networks: bacterial and yeast transcriptional networks and a mammalian signaling network contain far more sign-consistent feedback and feed-forward loops than expected for shuffled networks. Inconsistent loops with negative links can be more easily removed from real regulatory networks as compared to shuffled networks. This topological feature in real networks emerges from the presence of hubs that are enriched for either negative or positive links, and is not due to a preference for double negative links in paths. These observations indicate that intracellular regulatory networks may be close to monotone systems and that this network topology contributes to the dynamic stability.

## Introduction

Recently, three meso-scale (100-1,000 nodes) intracellular regulatory networks that specify the directionality and the effects of interactions have been developed: a mammalian cell signaling biochemical regulatory network in neurons (1), a bacterial (*Escherichia Coli*) gene regulatory network (2, 3), and a yeast (*Saccharomyces cerevisiae*) gene regulatory network (4, 5). These networks were constructed manually from the experimental literature and are derived from high-confidence functional experimental data. Knowledge of the functional consequence of the interaction allows the links to be characterized as positive for activation and negative for repression. Implicitly, the directions of interactions, i.e. the assignments of source and target nodes, are specified resulting in sign-specified directed networks. Since information flows through such networks, these can be considered dynamical systems.

In intracellular biochemical regulatory systems, the nodes can be proteins, metabolites, or genes, and the links can represent their direct interactions and/or indirect functional activity; for example: enzymatic, binding or translocation ability, or changes in overall quantity of the active form of a protein, or the quantity of a diffusible metabolite. Mathematically, sign-specified directed networks containing only “sign-consistent” loops, such as feed-forward or feedback loops (Fig. 1a), when converted to systems of differential equations, which represent the time evolutions of concentrations, always behave as monotone systems (6-8). Dynamical systems are labeled as monotone when a partial order of variables is preserved during time-progression of dynamical behavior (9). Monotone dynamical systems, extensively studied in control theory, are mathematically guaranteed to evolve in a predictable manner. They do not exhibit observable chaotic

behavior, while variable quantitative levels generically approach steady states as a function of time (9-12). Such dynamical behavior is commonly observed in cells. For example, bi-stability, multi-stability and monotone dynamics are typical in cell signaling regulatory networks and transcription quantitative levels in gene regulatory networks (13-16).

In this study we address the question of whether intracellular biological regulatory networks are close to monotone systems by analyzing three sign-specified directed networks. We assess the “distance to monotone” architecture by analyzing the level of “sign-consistency” in feedback and feed-forward loops identified in the topology of these networks. If these networks have a relative abundance of “sign-consistent” loops, this may explain observed intracellular

dynamics stability and order. We developed algorithms to characterize the abundance of positive “sign-consistent” (Fig. 1a) and negative “sign-inconsistent” feedback and feed-forward loops (Fig. 1b) in the three intracellular regulatory networks. We find that positive feedback and feed-forward loops are enriched in all three networks. This observation supports the hypothesis that one of us (ES) had made that biological intracellular regulatory networks may be close to monotone systems (12). We also find that abundance in positive feedback and feed-forward loops may be due to the enrichment of negative hubs and not because pathways contain disproportionate even number of negative links.

## Results

All three networks have similar nodes to links ratio, positive to negative links ratio, and display “small-world” properties (high clustering coefficients and similar characteristic path lengths compared to random networks) (Table 1). Counting the number of positive vs. negative feedback and feed-forward loops in the real regulatory networks vs. shuffled networks shows that there are significantly more positive loops than expected (Table 2). An approximate Binomial distribution analysis of the results is provided in Box 1. The shuffled networks used as a statistical control maintain the exact connectivity but differ in the distribution of signs (effects) associated with the links (random-swap), or in the assignment signs to links (positive vs. negative with probability  $p=0.5$ ) (random-sign) (see Methods). Interestingly, the difference between the real and shuffled for the yeast network is less significant with the random swap methods than the difference for the signaling and *E. coli*. This can be explained by the fact that feed-forward loops in the yeast network are highly nested. For example, during the procedure of removing the link that contributes to the most negative loops, the positive link between DAL80 and GLN3 caused the abolishment of 21 negative loops out of a total of 50. Both genes are GATA family transcription factors where DAL80 is an outgoing hub repressor regulated positively by GLN3 (17). The GATA family of genes makes up a complicated regulatory circuit which includes many members of the family regulating one another (18). Hence, the mathematical derivation in Box 1 assumes statistical independence of link contribution to loops, but in real network topologies the nesting of loops can drastically affect the distribution of positive to negative loops ratio also in shuffled networks, by making contribution of links to the formation of loops non-uniform, because some links may be reused to form many nested loops.

We developed an algorithm to remove links that contribute to the formation of inconsistent feedback and feed-forward loops. The algorithm, as demonstrated on a toy network (Fig. 2), gradually eliminates all small size negative loops (3-4 or 3-5 nodes per loop) from the real networks and from the shuffled networks. We find that it takes about one third (*E. coli* transcription), or one half (yeast transcription), or two thirds (mammalian signaling) of the links need to be removed from the real networks as compared to the number of links that need to be removed from corresponding shuffled networks (Fig. 3a, b, and c). This indicates that it is easier to convert the real networks to monotone “sign-consistent” topology as compared to shuffled networks.

The relative abundance of positive loops in the real networks, and the relative ease in removing the negative loops, could be due to either hub nodes that have many in- or out-going negative links, or because pathways in feedback and feed-forward loops tend to have double negative links in them, making loops to be considered positive. The first case implies that negative links are concentrated within regions of the networks, and thus increase the likelihood for forming sign-consistent positive feed-forward and feedback loops. Alternatively, positive feedback and feed-forward loops are abundant because they contain an even number of negative links. To determine which of these scenarios is more likely in real networks, we first plotted the in-links vs. out-links difference on the x-axis and positive-links vs. negative-links difference on the y-axis for all nodes (Fig 4a, b, and c). The plots show the existence of hubs with abundance of negative links in all three real networks compared with shuffled network. In particular, the yeast and the *E. coli* transcriptional network had many more out-going hubs including negative hubs. The signaling network had both positive and negative in and out hubs that were diminished after shuffling. All three networks show preferential enrichment for hubs with either only positive or only negative links. Because negative links are concentrated in different parts of the network, around a few hubs, and are not evenly spread around like in the shuffled networks, the probability of forming negative loops is reduced. Hence, the existence of hubs enriched in positive or negative links in the topology of real networks leads to the preference for positive feedback and feed-forward loops.

It is possible that the abundance of positive feedback and feed-forward loops in the real networks is due to pathways being rich in double negatives (even number of negative links in paths). Hence, we assessed whether the real networks are enriched with even number of

negative links in paths more than expected as compared to corresponding shuffled networks. For this, we used an algorithm that “starts” at any random node and explores random directed paths while counting the number of negative signs along the way. Pathways in the real networks did not show a propensity to contain an even number of negative links compared with pathways in the shuffled networks which have an unbiased probability to contain an even or an odd number of negative links in random directed paths (Fig 5a, b and c). Hence, double negatives in paths are not likely to contribute to the formation of consistent feedback and feed-forward loops, in the real networks studied here.

## Discussion

Our study demonstrates that intracellular biological regulatory networks may be close to monotone systems due to selection for positive feedback and feed-forward loops and selection against negative feedback and feed-forward loops. Inconsistent (negative) feedback loops have been shown to be more prone to produce rich and complicated dynamics (6). Hence, the selection against them provides an explanation as to why stable or multi-stable dynamical behavior is commonly observed in cells, and oscillations are rare. Although negative feedback loops intuitively may seem to be important for cellular homeostasis, the topology observed shows that negative feedback loops are not common, and homeostasis is probably maintained mostly through less intricately regulated mechanisms such as degradation and unregulated deactivation such as dephosphorylation by constitutive unregulated phosphatases in signaling networks. Stable or multi-stable dynamical behavior has been frequently observed experimentally in cells (13-16). Besides dynamical stability, monotone system architecture is also advantageous for ordered behavior and predictability, and evolutionary modularity.

Monotone systems are predictable and display ordered behavior (7, 8). For example, when we increase or decrease the concentration of a node or the rate constants for the interaction represented by a link in a network containing only positive feed-forward and feedback loops (19) the output would increase with time and then may decay due to constitutive negative regulators. In contrast, changes in the initial concentrations or rate constants in a network containing inconsistent feedback or feed-forward loops can induce oscillations or other complex behavior. Thus, monotone topology preserves input/output relationships between distal components in the network, a feature commonly observed in cell signaling pathways.

Monotone architecture is also useful for evolutionary modularity by limiting the propagation of changes to stay local. Consider a mutation in the most upstream component in an ensemble of nested feed-forward loops. The mutation provides enhanced adaptation for one part of the network while preserving the qualitative functional behavior of the remaining parts of the ensemble. Hence, a monotone architecture can assist in the preservation of modularity through network evolution.

The human genome has been found to have more types and isoforms of protein kinases than protein phosphatases (20). Our current understanding indicates that generally protein kinases in a regulated manner, selectively activates their downstream targets (1) although there are well-known examples where phosphorylation inhibits the activity of the protein. In contrast, protein phosphatases such as PP2A generally inhibit their targets and are considered “house-keeping” enzymes due to the assumption that they are less regulated than protein kinases and have many more substrates thus making them outgoing negative hubs. These interaction characteristics of the protein kinases vs. phosphatases may be the reason why the mammalian signaling network is a close to monotone system.

In conclusion, we have found that three intracellular regulatory networks have an unexpected low abundance of negative feedback and feed-forward loops compared to the number of negative loops in the corresponding shuffled networks. We also found that links that contribute to negative feedback and feed-forward loops can be easily removed to make the networks sign-consistent. This network topology results from an enrichment of hubs with many negative links and not due to the selection for pathways with an even number of negative links.

We conclude that intracellular regulatory networks have evolved to be mostly “sign-consistent” and thus are close to monotone systems. The dynamic stability of the cell may in part be due to this observed topology of the regulatory networks.

## Methods

### *Network datasets analyzed*

Signal transduction network representing interactions in mammalian neurons was assembled from literature (1) and downloaded from:

<http://amp.pharm.mssm.edu/data/9.22.2004.sig>

Escherichia Coli transcriptional regulation dataset (3) was downloaded from:

[http://www.weizmann.ac.il/mcb/UriAlon/Network\\_motifs\\_in\\_coli/ColiNet-1.1/](http://www.weizmann.ac.il/mcb/UriAlon/Network_motifs_in_coli/ColiNet-1.1/)

Saccharomyces Cerevisiae gene regulatory dataset (4) was downloaded from:

<http://www.weizmann.ac.il/mcb/UriAlon/Papers/networkMotifs/yeastData.mat>

These networks are directed graphs with three types of links: activation, inhibition, and neutral (signaling) or dual regulation (gene regulation).

### *Counting positive and negative cycles in the networks*

A positive consistent feedback or feed-forward loop is defined as containing an even number of negative links or no negative links. A feedback loop or a feed-forward loop is negative if it is not positive (1, 12, 21, 22). A recursive algorithm that uses depth-first search was developed to count positive and negative feedback and feed-forward loops (21). The neutral links in the signaling network and bidirectional links in the gene regulatory networks have not been considered valid links when counting feedback and feed-forward loops. Neutral and bi-directional links are not abundant in all three networks (table 1) and considering these links as either negative or positive does not significantly affect our results. For a definition of neutral links in the signaling network see reference (21) and for a definition of bi-directional links in the gene regulatory networks see reference (4).

### *Removing links that contribute to negative loops*

The following protocol was used to eliminate negative feedback and feed-forward loops of up to a certain size:

1. Apply the algorithm described above to find all the negative feedback and feed-forward loops of a certain size.
2. Sort links based on number of times links participate in found negative loops.



3. Remove the link that contributes to the most number of found negative loops.
4. Repeat until there are no more negative loops of a certain size.

Fig. 2 illustrates this concept on a toy network model. This algorithm was applied because removing all negative loops is NP-hard, although approximation algorithms for this task have been developed (23).

#### *Creating shuffled networks*

Shuffled signed networks were created from the original networks for use as statistical controls. The algorithm used to create these sign-shuffled networks is briefly described. Signs of links are randomly shuffled by picking randomly a pair of links and swapping their signs repeatedly. The shuffled signed networks maintain the same connectivity and maintain the same ratio of negative to positive links as the original networks. Randomly assigned signs networks were created by randomly assigning a positive or negative sign with  $p=0.5$  to all directed links.

## Acknowledgements

This research was supported by NIH grant GM-54508 to RI and by NSF grant DMS-0614371 to ES. We would like to thank Drs. Azi Lipshtat, Gustavo Stolovitzky, and Guillermo Cecchi for useful discussions.

## References

1. Ma'ayan, A., Jenkins, S. L., Neves, S., Hasseldine, A., Grace, E., Dubin-Thaler, B., Eungdamrong, N. J., Weng, G., Ram, P. T., Rice, J. J., Kershenbaum, A., Stolovitzky, G. A., Blitzer, R. D. & Iyengar, R. (2005) *Science* **309**, 1078-1083.
2. Salgado, H., Gama-Castro, S., Peralta-Gil, M., Diaz-Peredo, E., Sanchez-Solano, F., Santos-Zavaleta, A., Martinez-Flores, I., Jimenez-Jacinto, V., Bonavides-Martinez, C., Segura-Salazar, J., Martinez-Antonio, A., Collado-Vides, J. (2006) *Nucleic Acids Res.* **34**, D394-397.
3. Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. (2002) *Nat. Gene.* **31**, 64-68.
4. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. & Alon, U. (2002) *Science* **298**, 824-827.
5. Costanzo, M. C., Hogan, J. D., Cusick, M. E., Davis, B. P., Fancher, A. M., et al. (2000) *Nucl. Acids Res.* **28**, 73-76.
6. Angeli, D., Ferrell, J. E., Jr. & Sontag, E. D. (2004) *PNAS* **101**, 1822-1827.
7. Sontag, E. D. (2006) *arXiv.org q-bio.MN/0612032*.
8. Sontag, E. D. (2006) *arXiv.org q-bio.MN/0612033*.
9. Angeli, D., Sontag, E. D. (2003) *IEEE Transactions on Automatic Control* **48**, 1684-1698.
10. Hirsch, M. (1983) *Contemporary Mathematics* **17**, 267-285.
11. Smith, H. L. (1995) *Monotone dynamical systems: An introduction to the theory of competitive and cooperative systems* (Amer. Math. Soc., Providence).
12. Sontag, E. D. (2004) *Systems Biology* **1**, 9-18.
13. Bhalla, U. S., Ram, P. T. & Iyengar, R. (2002) *Science* **297**, 1018-1023.
14. Bhalla and Ravi Iyengar, U. S. (1999) *Science* **283**, 381-387.
15. Markevich, N. I., Hoek, J. B. & Kholodenko, B. N. (2004) *J. Cell Biol.* **164**, 353-359.

16. Ferrell, J. E., Jr. & Machleder, E. M. (1998) *Science* **280**, 895-898.
17. Cunningham, T. S. & Cooper, T. G. (1991) *Mol Cell Biol.* **11**, 6205–6215.
18. Cunningham, T. S., Rai, R. & Cooper, T. G. (2000) *J. Bacteriol.* **182**, 6584-6591.
19. Mangan, S. & Alon, U. (2003) *PNAS* **100**, 11980-11985.
20. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., et al. (2001) *Science* **291**, 1304-1351.
21. Ma'ayan, A., Blitzer, R. D. & Iyengar, R. (2005) *Annual Review of Biophysics and Biomolecular Structure* **34**, 319-349.
22. Remy, E., Mosse, B., Chaouiya, C. & Thieffry, D. (2003) *Bioinformatics* **19**, 172-178.
23. DasGupta, B., Enciso, G. A., Sontag, E. & Zhang, Y. *Biosystems* **In Press**.
24. Watts, D. J. & Strogatz, S. H. (1998) *Nature* **393**, 440-442.

**Table 1: Characteristics of the three intracellular regulatory networks**

<b>Network</b>	<b>Nodes</b>	<b>Links</b>	<b>CC</b>	<b>CPL*</b>	<b>Positive links</b>	<b>Negative links</b>	<b>Neutral or Bidirectional links</b>
<b>E. coli Ver. 1.1 gene regulation</b>	418	519	0.086	4.848	321	172	26
<b>S. cerevisiae gene regulation</b>	690	1082	0.047	5.208	860	221	1
<b>CA1 neuron Signaling</b>	546	1259	0.107	4.219	690	306	263

\* computed for the largest connected island

Statistical measurements for the networks: number of nodes, number of links, clustering coefficients (CC) and characteristic path lengths (CPL) (24), positive, negative and neutral or bidirectional links.

**Table 2 Positive and negative motifs in three intracellular regulatory networks.**

<b>Network</b>	<b>Positive feedback and feed-forward loops in real networks</b>	<b>Negative feedback and feed-forward loops in real networks</b>	<b>Positive loops in 20 randomly swapped signs networks</b>	<b>Negative loops in 20 randomly swapped signs networks</b>	<b>Positive loops in 20 randomly assigned signs networks</b>	<b>Negative loops in 20 randomly assigned signs networks</b>
<b>E. coli Ver. 1.1 gene regulation</b> (cycles size 3-5)	35	6	20.6 ± 3.0	19.45 ± 2.84	21.75 ± 4.01	19.25 ± 4.01
<b>S. cerevisiae gene regulation</b> (cycles size 3-5)	115	50	92.72 ± 11.57	72.75 ± 11.09	87.75 ± 5.41	76.55 ± 5.41
<b>CA1 neuron Signaling</b> (cycles size 3-4)	475	245	276.35 ± 35.03	260.1 ± 20.46	359.2 ± 8.39	360.8 ± 8.39

Comparison between positive and negative feedback and feed-forward loops found in the original networks and in shuffled networks created from the original networks using the recipe described in the methods. The numbers in the shuffled networks columns are average ± standard deviation. The reason that the totals (positives + negatives) for the randomized signaling networks are not the same as the real networks is because neutral links were also shuffled. This affects the counts of feedback and feed-forward loops which do not contain neutral links.

### Box 1: Analytical explanation for the distribution of positive vs. negative loops in shuffled networks

If there are  $P$  positive and  $N$  negative links, and  $P$  and  $N$  are sufficiently large, the probability of picking, using a Bernoulli process, a negative link is:

$$(1) p(-) = N / (N + P), \text{ and a positive link: } p(+) = 1 - p(-)$$

We define  $p(k)$  as the probability that a feedback or feed-forward loop is positive, where  $k$  is the number of links and nodes making up the loop. A positive loop is defined as a loop with either all positive links or an even number of negative links. Thus, we have the following linear first-order recurrence:

$$(2) p(k + 1) = p(+)p(k) + (1 - p(+))(1 - p(k)) \text{ with } p(1) = p(+).$$

This recurrence has the solution:

$$(3) p(k) = [1 + (2p(+)-1)^k] / 2$$

Thus, for  $0 < p(+) < 1$ ,  $p(k)$  converges to 0.5, and for  $p(+) = 1$ ,  $p(k) = 1$  (all links are positive), and  $p(k)$  alternates between 0 and 1 if  $p(+) = 0$  (the network is made of only negative links).

For example, for  $k = 5$  and the E-coli transcriptional network, where we have 321 positive links, 172 negative links, and 26 neutral links (we count neutral links as positive) we have:

$$p(+) = 347 / (347 + 172) = \text{approximately } 0.67.$$

Therefore, using this simplified Bernoulli argument, the probability of getting a positive loop is:

$$p(5) = [1 + (0.34)^5] / 2 = 0.502$$

(For  $k = 4$ ,  $p = 0.507$ , and for  $k = 3$ ,  $p = 0.52$ ).

## Figures Legends

### Fig. 1 Examples of “sign-consistent” and “sign-inconsistent” feed-forward and feedback loops.

(a) Examples of “sign-consistent” positive loops in the form of two positive feed-forward loops and one positive feedback loop. If the input node A is mutated or its concentration increases or decrease the output node D and the overall qualitative circuit behavior would be predictable. (b) In contracts, examples of negative “sign-inconsistent” feed-forward loops and a feedback loop are shown, where the output node D could be either overall increasing or decreasing as a result of a change in the properties of node A. Green arrows represent activation. Red plungers represent inhibition.

### Fig. 2 Toy network to illustrate the algorithm that removes links that contribute to negative feed-forward loops.

First, the algorithm counts the number of times a link contributes to the formation of negative feedback and feed-forward loops. Then the link that contributes to the most number of negative loops is removed from the network. The link from A to D is removed because it participates in two negative feed-forward loops (more than all other links). After this link is removed there are no more negative loops left in the toy network.

### Fig. 3

#### Gradual removals of links that contribute to negative loops.

The number of positive and negative feedback and feed-forward loops of size 3-5 in the *S. cerevisiae* and *E. coli* networks, and 3-4 in the CA1 signaling network were counted. Then, the link that contributes to the most number of negative loops was removed. The process is repeated until the networks no longer have small-size negative feedback and feed-forward loops. The results for the real networks are compared to applying the same procedure on randomly shuffled networks created from the original networks. (a) *E.Coli* gene regulation network. (b) *S. Cerevisiae* gene regulation network. (c) CA1 neuronal cell signaling network.

### Fig. 4

Visualization of positive-negative and in-out hubs. All nodes in the networks where positioned in a 2D grid based on an x-axis location as the difference between the in and out links for each node, and based

on y-axis location as the difference between the positive and negative links for each node. The results are compared to a randomly selected shuffled network plotted the same way. (a) *E.Coli* gene regulation network. (b) *S.Cerevisiae* gene regulation network. (c) CA1 neuronal cell signaling network.

**Fig. 5**

Tracing random directed pathways sign-consistency. Starting from 10,000 randomly chosen nodes, directed pathways of variant length in a random direction where traced from the number of negative links. The counts of pathways with 0 to 5 negative links (black bars) are compared to average and standard deviation computed by applying the same procedure to 20 shuffled networks. (a) *E. coli* gene regulation network. (b) *S. cerevisiae* gene regulation network. (c) CA1 neuronal cell signaling network.



Fig. 1

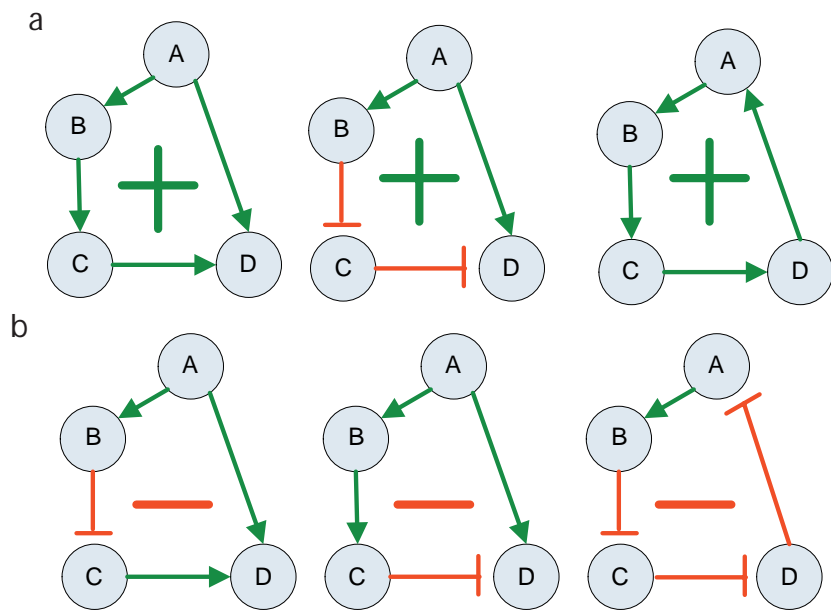


Fig. 2

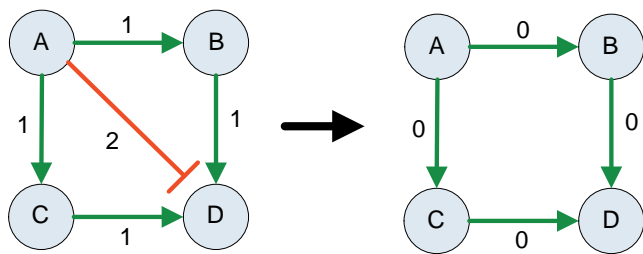


Fig. 3

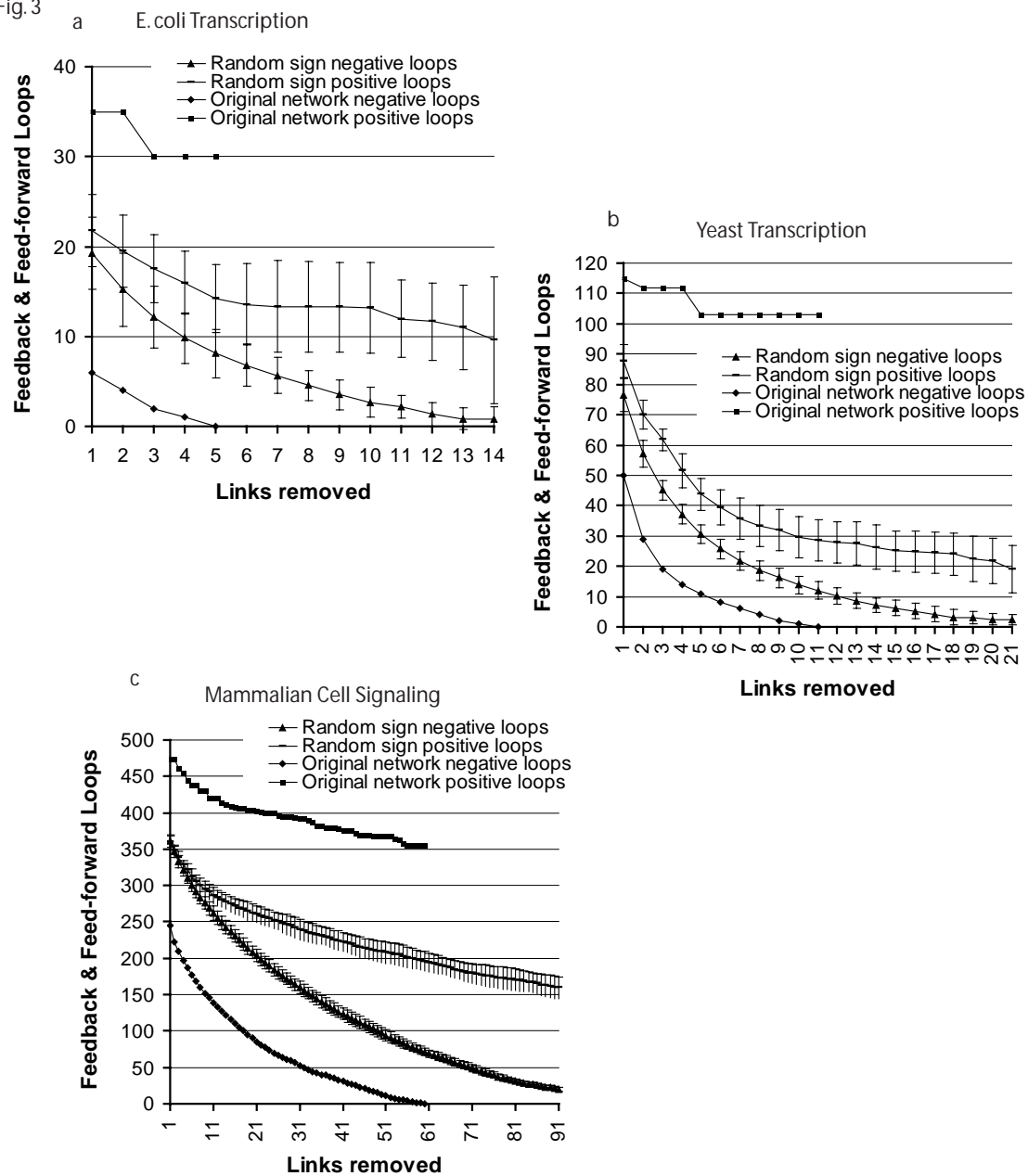


Fig. 4

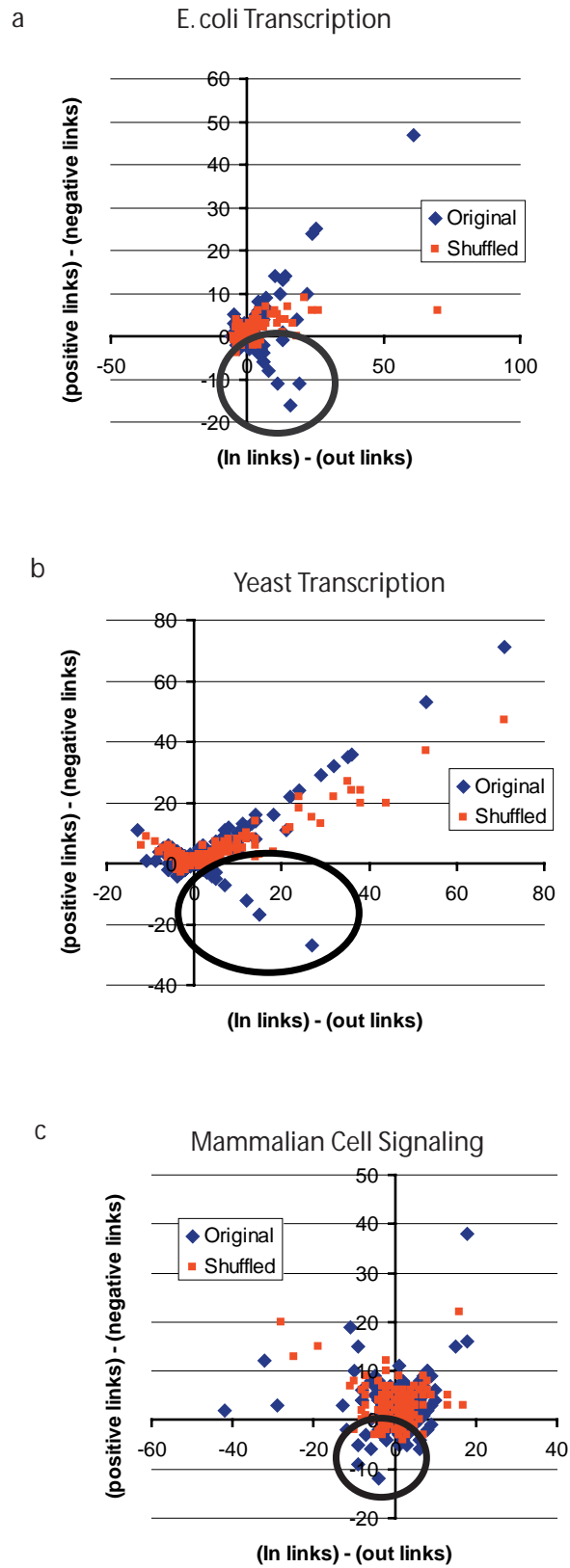


Fig. 5

