Evolutionary design principles and functional characteristics based on kingdom-specific network motifs

Tae-Hwan Kim, Junil Kim, Pat Hayes-Harrison and Kwang-Hyun Cho

1 Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Republic of Korea and 2 Department of Biology, University of Leicester, Leicester LE1 7RH, UK

ABSTRACT

Background: Network motifs within biological networks show non-random abundances in systems at different scales. Large directed protein networks at the cellular level are now well defined in several diverse species. We aimed to compare the nature of significantly observed two- and three-node network motifs across three different kingdoms (Arabidopsis thaliana for multicellular plants, Saccharomyces cerevisiae for unicellular fungi and Homo sapiens for animals).

Results: ‘Two-node feedback’ is the most significant motif in all three species. By considering the sign of each two-node feedback interaction, we examined the enrichment of the three types of two-node feedbacks [positive-positive (PP), negative-negative (NN) and positive-negative (PN)]. We found that PN is enriched in the network of A. thaliana, NN in the network of S. cerevisiae and PP and NN in the network of H. sapiens. Each feedback type has characteristic features of robustness, multistability and homeostasis.

Conclusions: We suggest that amplification of particular network motifs emerges from contrasting dynamical and topological properties of the motifs, reflects the evolutionary design principles selected by the characteristic behavior of each species and provides a signature pointing to their behavior and function.

Contact: ckh@kaist.ac.kr

Supplementary information: Supplementary data are available at Bioinformatics online.

Received and revised on September 2, 2010; accepted on October 25, 2010

1 INTRODUCTION

Systems biology involves the study of complex interactions between components of a biological system. The study of network motifs (Alon, 2007) within larger networks can help to unravel the functions of complex biological networks (Milo et al., 2002, 2004). In biology, network motifs—subgraphs that recur within a network much more often than expected in random networks—include negative autoregulation, positive autoregulation, feedforward loops, single-input modules, dense overlapping regulons and feedback loops (Alon, 2007). These network motifs are known to have different dynamical functions, and distinct subsets have been found to predominate in various biological networks such as transcription, signal transduction, protein interaction and neuronal networks, or even larger scale ecosystem-wide networks. Most of the previous studies on network motifs (Conant and Wagner, 2003; Prill et al., 2005; Wuchty et al., 2003; Yeger-Lotem et al., 2004) were, however, limited to the networks with no interaction sign (activation or inhibition) although such interaction sign is indispensable for characterizing various dynamical properties. Now that networks involving a substantial proportion of all the signed directed protein networks (SDPNs) within cells are available for organisms from the animal, plant and fungal kingdoms, we can compare the network motifs and ask whether there are characteristics which may be kingdom specific, and allow us to unravel evolutionary design principles. In this article, we examine the enrichment of particular network motifs in the Arabidopsis thaliana (multicellular plant), Saccharomyces cerevisiae (unicellular fungus) and Homo sapiens (multicellular animal). We suggest biological implications for the sets of network motifs in each species that arise from their characteristic dynamical and topological features.

2 METHODS

2.1 Construction of directed protein networks

The network data were retrieved from all available manually curated databases using Pathway Studio 6 (Nikitin et al., 2003). This software provides five biological network databases (ResNet Mammalian Database 6.0, ResNet Plant Database 2.0, Yeast-3, Drosophila-3 and C-elegans-3). Each database contains eight types of nodes (Small Molecule, Functional Class, Complex, Protein, Cell Process, Cell Object, Treatment and Disease) and eight types of links (Expression, Regulation, MolTransport, ProtModification, Binding, PromoterBinding, MolSynthesis and DirectRegulation). In order to obtain reliable directed protein interaction data, we retrieved only the referenced links comprised of ‘Protein’ nodes and five types of interaction (Regulation, MolTransport, ProtModification, MolSynthesis and DirectRegulation).

2.2 Identification of network motifs

In order to identify network motifs using mfinder (Kashtan et al., 2004; Milo et al., 2002), 1000 random networks were generated by the switching method that switches between edges while preserving the degree distribution of the nodes in the original network—in-degree, out-degree and mutual degree are preserved for each node. The number of switches was a random number in the range of 100–200 times the total number of edges in the network. Network motifs with $P<0.01$ was considered as statistically significant.
The degree of a node is a confounding factor in comparison of the topological average of the shortest paths from the node to all reachable upstream the average distance from a node to receptors (ADR) was defined as the average of the shortest paths from the node to all reachable downstream TFs. Similarly, average distance from a node to TFs (ADTF) was defined as the average of the shortest paths from the node to all reachable downstream TFs. Then, the average distance from a node to all reachable downstream TFs was calculated. Finally, a tissue-specific subnetwork (a tissue-specific subnetwork) was constructed by extracting the protein nodes in a tissue-specific protein set (a tissue-specific protein set) and their interactions from the SDPN of H.sapiens. For the comparison, 10 random subnetworks were also constructed by randomly extracting the same number of nodes with each tissue-specific/general subnetworks; hence, 200 random subnetworks were generated in total.

2.5 Average distance from a node to transcription factors or receptors

In the SDPNs of S.cerevisiae and H.sapiens, transcription factors (TFs) and receptors were selected as the nodes whose corresponding genes are classified into the Gene Ontology (GO) term ‘regulation of transcription (GO:0045436)’ and ‘receptor activity (GO:004872)’, respectively. According to this criterion, there are 503 TFs and 13 receptors for S.cerevisiae and 1894 TFs and 1354 receptors for H.sapiens. Then, the average distance from a node to TFs (ADTF) was defined as the average of the shortest paths from the node to all reachable downstream TFs. Similarly, the average distance from a node to receptors (ADR) was defined as the average of the shortest paths from the node to all reachable upstream receptors.

2.6 Degree-free sampling

The degree of a node is a confounding factor in comparison of the topological features of PP and NN–ADTF and ADR are inversely proportional to the degree of a node in a network. In order to standardize the topological features, we considered four degree intervals ([1, 5], [6, 10], [11, 20], [21, 30], [31, 50], [51, 100], and >100) for S.cerevisiae and seven degree intervals ([1, 5], [6, 10], [11, 20], [21, 30], [31, 50], [51, 100], and >100) for H.sapiens. In each mutually exclusive set of PP and NN, 100 nodes with degrees in each degree interval were randomly sampled with replacement; hence, 400 and 700 nodes were sampled for S.cerevisiae and H.sapiens, respectively. Then, we compared the average of each topological feature of PP with that of NN, and this procedure was repeated for 100 times.

2.7 GO analysis for two-node positive feedbacks

GO analysis enables us to determine which GO categories are statistically overrepresented in a set of genes or a subgraph of a biological network. In order to differentiate biological functions between the two-node positive feedbacks (PP and NN), we performed GO analysis by utilizing BinGO 2.31 (Maere et al., 2005), which is implemented as a plugin for Cytoscape (Shannon et al., 2003). Since we were interested in general functional profiling, a slimmed-down version of the whole GO (i.e. GO_Slim_GOA) was used for the analysis. Each mutually exclusive set of PP and NN was tested against its corresponding SDPN, and hypergeometric tests were performed for statistical analysis. P-values were corrected with Benjamini and Hochberg false discovery rate (FDR) and the significance level was set to 0.05.

3 RESULTS

3.1 Two-node feedback is the most significant network motif in the three different species

In constructing the biological networks of the three kingdoms (represented by A.thaliana, S.cerevisiae and H.sapiens), we retrieved directed protein interactions from all available databases (see Section 2.1; Fig. 1A). Directed protein interaction data for other model organisms such as Drosophila melanogaster and Caenorhabditis elegans also exist, but the network sizes are smaller, with many unknown interactions or more intensive study of particular pathways, and therefore are not sufficient for comparison with the three selected organisms (Supplementary Tables 1 and 2). Based on the networks, we performed network motif analysis (see Section 2.2). Significantly observed ($P < 0.01$) two- and three-node motifs are depicted as bold and colored structures in Figure 1B. The two-node feedback was the most significant network motif in all of the examined species (Fig. 1C and Supplementary Table 2). Notably, the number of significantly observed network motifs decreased for H.sapiens.

3.2 Two-node feedback motifs show enrichment characteristic of species: PN in A.thaliana, NN in S.cerevisiae and PP + NN in H.sapiens

To unravel the dynamical functions of two-node feedbacks, we examined the enrichment of three types of two-node feedbacks (PP, NN and PN) in each SDPN (Fig. 2A) where only signed interactions are extracted from the corresponding directed protein network (Fig. 1A). The $Z$-scores in Figure 2B (see Section 2.3; Supplementary Table 3) show that PN was enriched ($Z$-score $> 0$) only in A.thaliana, while in S.cerevisiae NN was significantly ($P < 0.01$) enriched and in H.sapiens both PP and NN were significantly enriched. To examine whether this enrichment characteristic depends on the particular network reconstruction method or the (possibly) incomplete network information we employed, randomly subsampled networks for each species were further investigated. As a result, we confirmed that the enrichment results are not biased by such a network reconstruction method or incomplete network information (see ‘Enrichment of two-node feedbacks in the sub-sampled networks’ in Supplementary Material and Supplementary Fig. 1).

Why is PN enriched only in A.thaliana? Negative feedback is important in homeostasis and robustness to perturbation (Brandman and Meyer, 2008; Brandman et al., 2007). Consistently, as we removed the protein nodes classified into the GO (The Gene Ontology Consortium, 2006) term ‘homeostatic process (GO:0042592)’, PN was no longer enriched ($Z$-score $< 0$;
Kingdom-specific network motifs for evolutionary design principles

Fig. 1. Network motifs in the three different species (A. thaliana, S. cerevisiae and H. sapiens). Interactions in each network were retrieved from Pathway Studio with nodes involved in two-node feedback motifs colored yellow. (A) Directed protein networks of A. thaliana (blue: 1750 nodes and 3705 links), S. cerevisiae (red: 1653 nodes and 3860 links) and H. sapiens (green: 8053 nodes and 72588 links). (B) Network motifs of the three different species. Significantly observed (P < 0.01) two- and three-node motifs are represented as bold and colored structures below each network topology. (C) Z-scores of significantly observed two- and three-node network motifs for the three different species. Note that the two-node feedback is the most significant network motif in each species.

Supplementary Table 4). PN enrichment was also found in the SDPN of rice, Oryza sativa (Supplementary Table 3). Moreover, some types of three- and four-node negative feedbacks (PPN, NNN and PPPN) were also enriched in A. thaliana although only PPPN was enriched in case of O. sativa (Supplementary Table 5). The contrast with fungi and animals suggests that negative feedback leading to robust homeostasis is selected in plants.

Like H. sapiens, two other multicellular animals (Mus musculus and Rattus norvegicus) also exhibited the significant enrichment of PP and NN (Supplementary Table 3). Species-specific differential enrichment of PP and NN suggests that the characteristics of each species might have emerged from different dynamical functions of PP and NN. Positive feedbacks are widespread in developmental processes as a result of their dynamical functions of hysteresis and multistationarity (Kim et al., 2008a): PP drives the developmental system towards a monostable regime (Chickarmane et al., 2006; Legewie et al., 2006), while NN leads to a bi- or multistable regime in development (Hwang et al., 2005; McClean et al., 2007), outcomes that have been noted as progression and decision switches (Guantes and Poyatos, 2008). Might alternative controls be favored in a single-cell organism such as S. cerevisiae where rapid changes in cell environment are encountered and most cells remain totipotent? Decision switches might be utilized in multipotent regimes in developmental processes. In contrast, a multicellular animal such as H. sapiens has many types of terminally differentiated cells in progressive differentiation processes, each protected from the environment. Hence, it might undergo developmental processes not only through decision switches but also through progression switches. Arguably, NN might act as a decision switch for bi- or multistable differentiation (e.g. from a stem cell to a pluripotent progenitor cell, from a pluripotent progenitor cell to a unipotent progenitor cell), whereas PP could act as a progression switch for monostable differentiation (e.g. from a unipotent progenitor cell to a terminally differentiated cell).

3.3 Coupling tendency of two-node feedbacks is different for each species

Biological networks involve many interlinked feedbacks (Brandman et al., 2005; Kim et al., 2007a, b, 2008b; Tsai et al., 2008), and we aimed to examine how the three types of two-node feedbacks are coupled in the species. Each SDPN (Fig. 2A) was converted
Fig. 2. Enrichment of two-node feedbacks and their couplings in the three different species (A.thaliana, S.cerevisiae and H.sapiens). (A) SDPNs of A.thaliana (1286 nodes and 1959 links), S.cerevisiae (1140 nodes and 1848 links) and H.sapiens (6958 nodes and 47469 links). (B) Enrichment of the three types of two-node feedbacks in each SDPN. Filled bars represent the significantly enriched (P < 0.01) network motifs. (C) Representation of the three types of two-node feedbacks (PP, NN and PN) and the six types of two-node feedback couplings (PP–PP, NN–NN, PN–PN, PP–NN, NN–PN and PN–PP). Blue arrows and red arrows with blunted ends represent positive and negative regulations, respectively. (D) Two-node feedback coupling networks of A.thaliana (102 nodes and 185 links), S.cerevisiae (76 nodes and 116 links) and H.sapiens (1631 nodes and 19213 links). (E) Enrichment of the six types of two-node feedback couplings in each two-node feedback coupling network. Filled bars represent the significantly enriched (P < 0.01) network motifs.

into a two-node feedback coupling network (Fig. 2C and D) where each node denotes one of the three types of two-node feedbacks (PP: blue, NN: red, PN: green) and each link denotes a coupling, or a node sharing between two two-node feedbacks (right box of Fig. 2C). Statistical analysis in Figure 2E (see Section 2.3; Supplementary Table 3) shows that PN–PP was relatively enriched in the network of A.thaliana. Only homogeneous couplings (PP–PP, NN–NN and PN–PN) were significantly enriched in the network of S.cerevisiae. In the case of H.sapiens, PP–PP, PN–PN and PN–PP were significantly enriched in the network. These findings and the results from other species (Supplementary Table 3) suggest the following three points: (i) homogeneous couplings are more enriched in the networks of single-cell organisms (S.cerevisiae) and multicellular animals (H.sapiens, M.musculus and R.norvegicus) than heterogeneous couplings (PP–NN, NN–PN and PN–PP), (ii) PP–PP is enriched in all the species, whereas PP–NN is not enriched in any of the species and (iii) for heterogeneous couplings, PN tends to be coupled with PP in the networks of multicellular organisms, whereas PN tends to be coupled with NN in the networks of single-cell organisms.

3.4 Dynamical features of two-node feedbacks

We noted above that species-specific differential enrichment of two-node feedbacks could originate from their different functions. To validate this, we investigated the dynamical features of PP, NN and PN, first by evolving artificial networks with a preference for ‘multistationarity’ (Fig. 3A), ‘robustness of point attractor’ (Fig. 3B) or ‘robustness of cyclic attractor’ (Fig. 3C) using Boolean network simulations (Kim et al., 2008a) (see ‘Network evolution’ in Supplementary Material). The artificial evolution toward ‘multistationarity’ increased the numbers of two-node positive feedbacks while suppressing negative feedbacks (Fig. 3A and D). We note that NN becomes more preferential than PP as the number of nodes in artificial networks increases (Supplementary Fig. 2 and Supplementary Table 6). The preference of NN...
Kingdom-specific network motifs for evolutionary design principles

Fig. 3. Dynamical features of the two-node feedbacks. (A–C) The average numbers of two-node feedbacks during the evolution of four-node artificial networks with a preference for multistationarity (A), robustness of point attractor (B), or robustness of cyclic attractor (C). (D–F) Attractor landscape of a network (inset figure) having the best fitness score for multistationarity (D), robustness of point attractor (E) or robustness of cyclic attractor (F). Attraction fields for every possible state of given networks are mapped to the corresponding attractor landscapes. The z-axis in the attractor landscape denotes the energy potential measured by a distance to the corresponding attractor.

over PP, although both are two-node positive feedbacks, might be because PP-only couplings can incorporate only two 'point attractors' whereas NN-only couplings can give rise to multiple 'point attractors' (Kim et al., 2008a). When artificial networks were evolved to enhance ‘robustness of point attractor’, PN and PP interactions were increased whereas NN was decreased (Fig. 3B and E). In the case of ‘robustness of cyclic attractor’, only PN was increased while suppressing positive feedbacks (Fig. 3C and F).

In summary, we found that NN is exploited only for multistationarity, PP increases both multistationarity and robustness of point attractor, and PN enhances robustness of attractors. These results support our previous findings that PP can be employed for a unipotent regime as a discreet decision maker, NN for a multipotent regime for multistability—multistationarity is essential for multistable differentiation (Laslo et al., 2006)—and PN for robust homeostatic behavior.

3.5 Topological features of two-node positive feedbacks

From the different dynamical roles of two-node positive feedbacks in developmental processes, we speculated that if PPs (NNs) are exploited for unipotent (multipotent) regimes, genes comprising PP (NN) structure would be expressed in specific (diverse) conditions for single-cell organisms and in specific (diverse) tissues for multicellular organisms. To test this idea, we compared the responsiveness, defined as the sum of squares of the log2-ratios measured from curated expression datasets under more than 1500 conditions (Choi and Kim, 2009; Tirosh et al., 2006), of PP versus NN and PP–PP versus NN–NN in S. cerevisiae. We found that NN and NN–NN show higher responsiveness to various conditions than PP and PP–PP, respectively (Fig. 4A). This result implies

Fig. 4. Topological features of the two-node positive feedbacks (see Section 2). (A) Responsiveness of the mutually exclusive sets of PP versus NN and PP–PP versus NN–NN in S. cerevisiae. (B) Enrichment of PP! NN in tissue-specific and tissue-general subnetworks for H. sapiens. (C, D) Scatter plots for the average ADTFs and ADRs for the mutually exclusive sets of PP and NN in S. cerevisiae and H. sapiens. In (B–D), red data points indicate mean values for each data category.
Moreover, PPs were relatively closer to the receptors than NNs closer to the TFs in the SDPN of Figure 4C, PPs were closer to the receptors, while NNs were involved in exogenous (endogenous) processes. Exogenous processes are a means to cope with the cell environment through receptor-mediated signal transduction, whereas endogenous processes are employed for various cellular decisions by utilizing endogenous processes.

Species | GO-terms for PP | GO-terms for NN
---|---|---
*S. cerevisiae* | Cell communication\(^a\) | Regulation of biological process\(^b\)
 | Signal transducer activity\(^b\) | Nucleus\(^b\)
 | Membrane | Enzyme regulator activity\(^b\)
*H. sapiens* | Cell communication\(^a\) | Regulation of biological process
 | Signal transducer activity\(^a\) | Binding
 | Response to stimulus\(^a\) | Protein binding
 | Kinase activity | Cell motility
 | Cellular process | Extracellular region\(^a\)
 | Behavior | Transmembrane transport activity
 | Binding | Cell death
 | Extracellular space\(^a\) | Molecular function
 | Receptor activity\(^a\) | Macromolecule metabolic process
 | Biological process | Transporter activity
 | Extracellular region\(^a\) | Cell death
 | Transmembrane transport activity | Cell death

Only the GO terms with corrected P < 0.05 are listed (see Supplementary Tables 7–10 for more details).

\(^a\)GO terms related to exogenous processes.

\(^b\)GO terms related to endogenous processes.

that in adaptation to rapidly changing environment, NNs are exploited for totipotent regimes leading to different cell types in *S. cerevisiae*. For *H. sapiens*, we examined the enrichment of PP and NN in 10 tissue-specific (filled green circles in Fig. 4B) and 10 tissue-general subnetworks (filled blue triangles in Fig. 4B), comparing with 10 random subnetworks for each (empty sky blue circles for tissue-specific and empty orange triangles for tissue-general subnetworks in Fig. 4B). As a result, the tissue-specific (tissue-general) subnetworks showed higher enrichment for PP (NN) than the random subnetworks, implying that PPs (NNs) are utilized for unipotent (multipotent) regimes leading to fully differentiated cells (various cell types) in *H. sapiens*.

In order to further understand the different biological aspects of two-node positive feedbacks, we performed GO analysis (see Section 2.7). According to the result shown in Table 1, PPs (NNs) are involved in endogenous (exogenous) processes. Exogenous processes are a means to cope with the cell environment through receptor-mediated signal transduction, whereas endogenous processes are employed for various cellular decisions by utilizing transcriptional programs in the nucleus. In support of this, as shown in Figure 4C, PPs were closer to the receptors, while NNs were closer to the TFs in the SDPN of *S. cerevisiae* (see Section 2.5). Moreover, PPs were relatively closer to the receptors than NNs were, although both PPs and NNs were closer to the receptors than TFs in *H. sapiens* (Fig. 4D). In multicellular animals, cells need to communicate in order to act in concert with other cells, thereby increasing the importance of exogenous processes such as cell-to-cell communications as indicated by the increased proportion of PP interactions in *H. sapiens* (Fig. 2B).

### 4 CONCLUSIONS

Network motifs have particular functional importance in various biological networks, and our results show that different classes of motif are enriched in different species from the three kingdoms. The data are still limited because network topology is only available for an adequate number of nodes in three species, and even so there may be ascertainment biases because the networks do not account for all cellular protein networks. However, for the first time we reveal species-specific network motif signatures that implicate contrasting evolutionary driving forces. In summary, (i) PN enrichment only in *A. thaliana* represents the importance of homeostasis in plants, (ii) NN enrichment in *S. cerevisiae* indicates the prevalence of multipotency in the developmental processes of this unicellular fungus and (iii) PP and NN enrichment in *H. sapiens* represents the progressive and often largely irreversible nature of the developmental processes. As further data are collected, it will be critical to contrast different abundance of network motifs between kingdoms with unicellular and multicellular forms, and to see whether the most abundant motifs are more similar in network regions concerned with similar functions but without conserved pathways, such as resistance to biotic or abiotic stresses. Network motif analysis may provide a valuable numerical tool to represent and compare large cellular protein networks between species, giving a signature that is a pointer to their function, behavior and evolution.

### ACKNOWLEDGEMENTS

The authors would like to thank Jung Kyoon Choi and Philip Murray for their helpful discussions.

Funding: National Research Foundation of Korea (NRF) grants funded by the Korea Government, the Ministry of Education, Science & Technology (MEST) (2009-0086964 and 2010-0017662).

Conflict of Interest: none declared.

### REFERENCES


Brandman, O. et al. (2007) STIM2 is a feedback regulator that stabilizes basal cytosolic Ca\(^{2+}\) levels. *Cell*, 131, 1327–1339.


Choi, J. K. and Kim, Y. J. (2009) Intrinsic variability of gene expression encoded in the developmental processes. As further data are collected, it will be critical to contrast different abundance of network motifs between kingdoms with unicellular and multicellular forms, and to see whether the most abundant motifs are more similar in network regions concerned with similar functions but without conserved pathways, such as resistance to biotic or abiotic stresses. Network motif analysis may provide a valuable numerical tool to represent and compare large cellular protein networks between species, giving a signature that is a pointer to their function, behavior and evolution.

### ACKNOWLEDGEMENTS

The authors would like to thank Jung Kyoon Choi and Philip Murray for their helpful discussions.

Funding: National Research Foundation of Korea (NRF) grants funded by the Korea Government, the Ministry of Education, Science & Technology (MEST) (2009-0086964 and 2010-0017662).

Conflict of Interest: none declared.

### REFERENCES


Brandman, O. et al. (2007) STIM2 is a feedback regulator that stabilizes basal cytosolic Ca\(^{2+}\) levels. *Cell*, 131, 1327–1339.


Choi, J. K. and Kim, Y. J. (2009) Intrinsic variability of gene expression encoded in the developmental processes. As further data are collected, it will be critical to contrast different abundance of network motifs between kingdoms with unicellular and multicellular forms, and to see whether the most abundant motifs are more similar in network regions concerned with similar functions but without conserved pathways, such as resistance to biotic or abiotic stresses. Network motif analysis may provide a valuable numerical tool to represent and compare large cellular protein networks between species, giving a signature that is a pointer to their function, behavior and evolution.

### ACKNOWLEDGEMENTS

The authors would like to thank Jung Kyoon Choi and Philip Murray for their helpful discussions.
Kingdom-specific network motifs for evolutionary design principles