BMC
Systems Biology

On the architecture of cell regulation networks

Yueheng Lan[1], Igor Mezić[2,3]*

Abstract

Background: With the rapid development of high-throughput experiments, detecting functional modules has become increasingly important in analyzing biological networks. However, the growing size and complexity of these networks preclude structural breaking in terms of simplest units. We propose a novel graph theoretic decomposition scheme combined with dynamics consideration for probing the architecture of complex biological networks.

Results: Our approach allows us to identify two structurally important components: the "minimal production unit"(MPU) which responds quickly and robustly to external signals, and the feedback controllers which adjust the output of the MPU to desired values usually at a larger time scale. The successful application of our technique to several of the most common cell regulation networks indicates that such architectural feature could be universal. Detailed illustration and discussion are made to explain the network structures and how they are tied to biological functions.

Conclusions: The proposed scheme may be potentially applied to various large-scale cell regulation networks to identify functional modules that play essential roles and thus provide handles for analyzing and understanding cell activity from basic biochemical processes.

Background

Cellular behavior, including motility, metabolism and reproduction is controlled by complex biochemical reaction networks, many of which have been identified and studied in detail [1]. These networks realize their regulatory roles through complex molecular interactions. Contemporary high throughput experiments produce unprecedented amount of data that serve to pinpoint the players and their interactions, resulting in complex chemical reaction graphs. How to analyze these intricate graphs and gain insight into the regulation mechanism employed by cell has become a central problem of molecular biology.

Much progress has been made in the analysis of functions of complex networks, no matter if they are modeled deterministically [2,3] or stochastically [4-9]. These studies concentrate on the investigation of dynamics of given networks by checking their stability, parameter dependence, robustness and input-output relation. However, for large-scale networks such as those commonly found in important biological processes [10,11], the incurred computational load often severely limits our ability for performing detailed analysis. More critically, with continued experimental efforts that are revealing more details of networks' global wiring, their growing complexity has made it harder and harder to identify the underlying local functional structures and thus probe the network function.

Normal cell life involves physical or chemical activities at vast range of spatial and temporal scales and it is vital to identify characteristic structures at all scales and study their roles in relation to a particular cell function [12-17]. These key structures are called modules, the existence of which contributes almost to every aspect of the cell regulation: robustness, sensitivity, adaptivity, evolvability. Their detection and study much simplifies the analysis of complex networks since a small set of modules could come from and be a lot simpler than a collection of many entangled individual agents [18]. The simplification may be carried on by constructing modules of modules.

Recently, useful concepts distilled from statistical physics such as the small-world and the scale-free networks [19,20], began to see their application in gene regulation

---

* Correspondence: mezic@engineering.ucsb.edu
[2]The Center for Control, Dynamical Systems and Computation, University of California, Santa Barbara, CA 93106, USA
Full list of author information is available at the end of the article

networks and lead to considerable success in unraveling the statistical nature of these networks. However, this type of statistical analysis mainly aims at gross features of networks [21] and thus ignores local structural properties and heterogeneities, which often determine the operation of a network in an essential way, since disparate network modules generally imply distinct dynamics and fit for different functional requirements [22,23]. Nevertheless, the determination of modular structure in a large network is not straightforward since one molecular species may be involved in many different pathways with very distinct external connections. Such inter-correlation is easily under-appreciated and yet has profound consequences on the organism.

In this paper we propose a new theory of architecture of biochemical networks based on control and graph theoretic analysis. In this theory, a network consists of two major modules: one is the pipeline of linear information production unit which serves to generate the required output (e.g. protein concentrations); the other is the set of feedback loops which act as controllers of the production. These two modules are identified based on the information flow in a network. Specifically, input and output nodes define a polarity of the network. Information is received at the input, processed and then sent to the output. The agents that carry on the information along the forward direction belong to the production unit. The remaining agents direct part of the information in the opposite direction and thus are elements of the feedback controller [22]. In the paper, detailed algorithm are presented for the construction of the production unit and the feedback controller.

The concept of modules has been used in modeling of biological networks for decades. The existence of this special structure is universally agreed upon but its exact definition is done on case-by-case basis. Recently, modules and community structures are defined in the graph theoretic studies of many real-world networks [20,24], based on the connectivity between nodes. Useful as it is, this type of definitions ignore the importance of controller loops. The community structure in the synchronization study involves more dynamics information but it works for a special class of networks and for particular types of equations of motion. Closely related concepts, such as "network motif" are also proposed [13,25]. Motifs consist of a small number of nodes and appear repeatedly (more than expected from pure statistical consideration) in a network. The modules determined by our algorithm are different from all these in that we emphasize the information processing and controlling units but not simple fixed graph structures given a priori. In contrast, the decomposition procedure based on the function of the network and the associated polarity supplies the detailed structures of our modules.

Different polarities may result in different decompositions and different initial conditions may define different MPUs. So our concept of modules depends on the information flow through or the function of the network.

In the following, we will use the NFκB regulation network [26] as an example to explain our graph theoretic analysis procedure and display the generic producer-controller structure. We also analyze the chemotaxis network of *E. coli*, TNF-α initiated apoptosis network [27], the circadian clock network in *Drosophila* [28] as interesting examples of the proposed architecture. Three more examples of biological networks are presented in Additional File 1 and are all found to possess the same architecture.
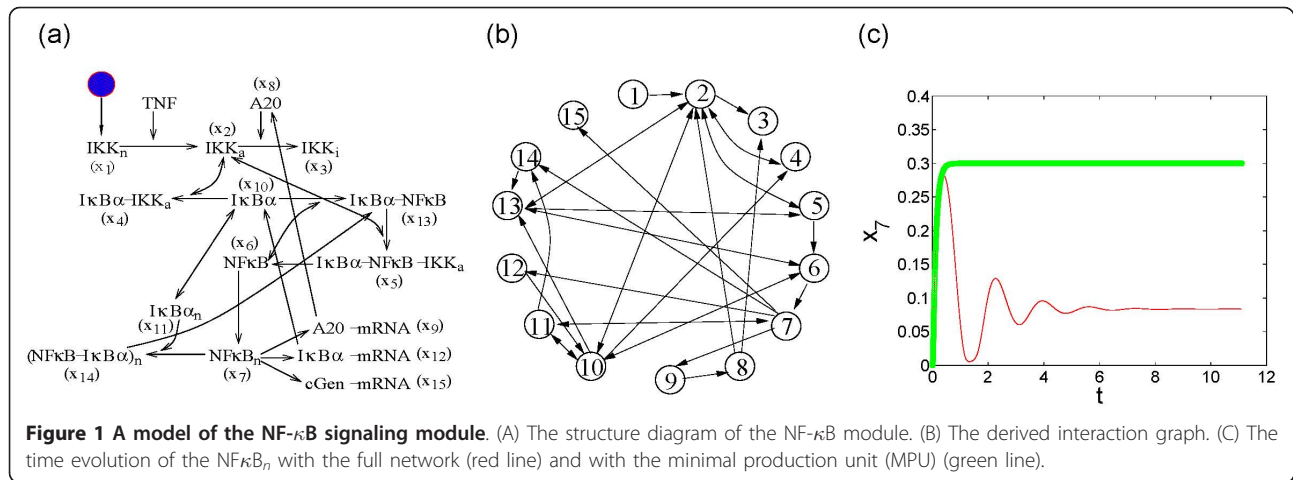
Results and Discussion
The NFκB regulatory network

The NFκB regulatory pathway concerns the switching dynamics of the nuclear factor NFκB, which regulates various genes important for pathogen or cytokine inflammation, immune response, cell proliferation and survival [29,30]. In the cytoplasm of a resting cell, NFκB usually binds to IκBα and its activity is suppressed. Certain external signals activate the switch protein IKK which phosphorylates IκBα such that NFκB is released [31]. The free NFκB then translocates into the nucleus and initiates the transcription of a large set of proteins, including protein IκBα and protein A20. Protein IκBα, once synthesized in the cytoplasm, enters the nucleus, binds to NFκB, transports it out to the cytoplasm and thus terminates the transcription. Protein A20 deactivates IKK. Therefore, the module mainly consists of two forward proteins IKK and NFκB and two feedback proteins IκBα and A20. Also, the translocation of the proteins between the nucleus and the cytoplasm is an important biological process that realizes spatial localization of different protein species.

The diagram of a detailed model of the NFκB regulatory network is shown in Figure 1A where we use x_i's to represent the concentration of various proteins. The associated chemical kinetic model is given and explained in Additional File 1. With physiological initial conditions [32], the concentration of the nuclear NFκB executes damped oscillations, shown with the thin dotted curve in Figure 1C. At the beginning, it shoots up to a very high value in a short time and then relaxes to a much lower steady value in an oscillatory way.

For any networked system described by certain dynamical equations, it is easy to write an interaction graph with the vertices representing the reacting agents and the edges directed from each agent to the ones under its influence. The interaction graph for the NFκB model is shown in Figure 1B.

Figure 1 A model of the NF-κB signaling module. (A) The structure diagram of the NF-κB module. (B) The derived interaction graph. (C) The time evolution of the NFκB$_n$ with the full network (red line) and with the minimal production unit (MPU) (green line).

It is straightforward to write down the adjacency matrix for the interaction graph, which marks 1 at the entries corresponding to connected edges and zero otherwise. The interaction graph and the adjacency matrix neglect details of the interactions and only map out the network topology which holds true almost everywhere in the phase space and the parameter space, except for a set of measure zero [33]. This robustness confers flexibility of analysis to analyzing vastly different dynamics described by ODEs or mappings or even stochastic equations. Certain system properties, like the uniqueness of the stationary point sometimes can be deduced from pure topological consideration of network structures [34,35]. So, understanding of structure of interaction graphs helps unveil the key elements in a complex system which possibly has uncertainties in the parameter values or is influenced by a noisy environment. Graph theoretic techniques will be developed here to enable an automatic decomposition of a biochemical network into forward and feedback modules, thus unraveling the architecture responsible for its biological function.

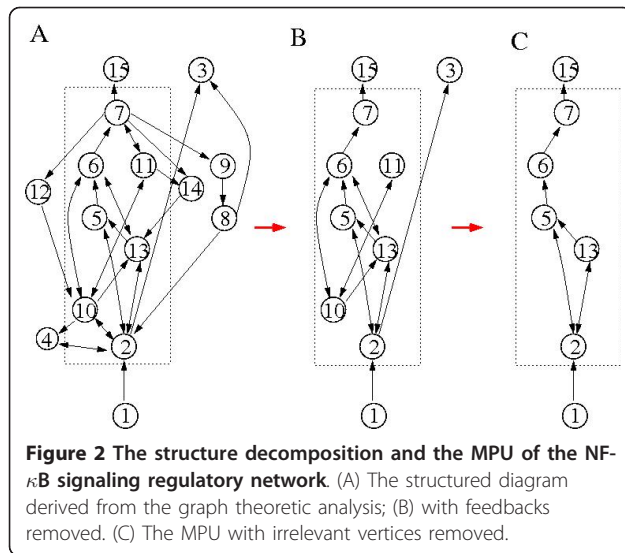**Controllers of the NFκB network**

The horizontal-vertical decomposition (HVD) of an interaction graph of a dynamical system has been discussed in a paper [33]. It is a technique that studies information flow and processing in interconnected systems. Vertically, the HVD decomposes a system into a linear series of layers, where the layer downstream is influenced by the layer upstream but not vice versa. So, the input signal propagates unidirectionally. Horizontally, the HVD decomposes each layer into independent groups with no direct connections between. In one layer, each group receives its own input from upstream layers and output the signal to downstream layers. Each group is a strongly connected component (SCC) such

that a path always exists between any two nodes in the group. If each group collapsed into a point, the whole network will become cycle-free [36].

Direct application of the HVD to the interaction graph in Figure 1B results in three layers with the top and bottom layer consist of the vertex sets $\{x_1\}$(IKK$_n$) and $\{x_3, x_{15}\}$(IKK$_i$, cGen-mRNA), respectively. The rest of the vertices are strongly connected and belong to the middle layer. This type of structure with dominant intermediate processing unit exists in most biological and engineering networks [33,37] as a result of omnipresent feedback loops and reversibility of many biochemical reactions. Below, we apply our cycle search and selection technique to the middle layer for further decomposition into the production unit and feedback controller.

The polarity of the middle layer is ready to be identified. The vertex x_{15} is the output signal that is of interest while x_1 receives the external input. Therefore, in the middle layer, x_2 is the input vertex and x_7 is the output one. In the mean time, we observe that in an SCC, if feedbacks exist, they are always making cycles and vice versa every cycle contains at least one forward and one feedback edge. As cycles are obvious topological invariants of a network and easy to seek, our strategy consists of two steps: first, search for all cycles that exist in the graph; second, determine the feedbacks through a selection procedure, which depends on the polarity of the network. The detailed illustration of our technique is contained in the Methods section. Here we show the computation result in Figure 2A, where we see that our procedure identified four feedback loops:

• FB$_a$ - the one through vertex 4: IKKa associates with free IκBα and catalyzes its decay.
• FB$_b$ - the one through vertex 14: IκBα_n captures NFκB$_n$ to form (IκBα-NFκB)$_n$, which then moves out of the nucleus.

Figure 2 The structure decomposition and the MPU of the NF-κB signaling regulatory network. (A) The structured diagram derived from the graph theoretic analysis; (B) with feedbacks removed. (C) The MPU with irrelevant vertices removed.

• FB_c - the one through vertex 12: $NF\kappa B_n$ promotes the production of the $I\kappa B\alpha$ mRNA which translocates to the cytoplasm and initiates a burst of $I\kappa B\alpha$ production.

• FB_d - the one through vertices 8 and 9: $NF\kappa B_n$ promotes the production of the A20 mRNA and thus initiates the production of A20, which catalyzes the decay of IKK_a.

This identification agrees very well with the usual recognition of feedback loops of this system in the literature [29,30] based on biological reasoning. The correct identification of feedback loops is essential for understanding the signal processing of a network since many important cellular activities are controlled or even realized by feedback signaling [22,23]. We emphasize that we recognized the feedback loops by an automatic procedure based on graph decomposition.

Extracting the minimal production unit

After the structured network is constructed as in Figure 2A, we proceed to the extraction of the minimal production unit (MPU). In the case of signal transduction network, the MPU is the minimal subgraph of a network that produces a response to external stimuli. The MPU is minimal in the sense that removal of any links from the subnetwork will lead to zero output. However, the response of the MPU may happen at a value that is different from what is desired in a real cell and setting that correct value is one of the roles of the feedbacks. Its identification depends both on the initial state of the system and on the signal that is of interest. Moreover, certain qualitative aspects of chemical kinetics of the network need to be considered in the course. As a matter of fact, the binary or dissociative reactions correlate

certain edges that represent same reactions. For example, the associative reaction $A + B \rightarrow C$ is depicted as $A \rightarrow C \leftarrow B$ in the interaction graph and the two arrows represent the same reaction. In previous computation, we ignored this correlation and carried out our analysis purely from a graph theoretic point of view. A more detailed consideration needs to incorporate this correlation: these two arrows have to coexist. Below, the NFκB network is used as an example to demonstrate the procedure of the MPU extraction in detail.

As we now only consider the forward production part to output x_{15}, the feedbacks and the associated reactions are first removed. For the NFκB network, we remove $\{x_4, x_8, x_9, x_{12}, x_{14}\}$ and arrive at Figure 2B. The correlation among edges has been considered as suggested by the above-mentioned binary reaction, i.e., the correlated arrows will be removed or kept coincidentally. Next, all the outputs except the one we are interested in are removed. That is, $\{x_3, x_{11}\}$ are removed. Here we see that the final MPU indeed depends on what signal we are looking at. Different output may result in different MPUs. Finally, we remove other irrelevant vertices in a recursive way according to the topology of the resulting graph and the given initial conditions. In the NFκB example, based on Figure 2B, x_{10} is removable since it does not lie on the main information path and $x_{10}(t) = 0$ all the time with $x_{10}(0) = 0$ being given. All this being done, we produce the MPU depicted in Figure 2C.

The MPU of the NFκB network contains the vertex set $S_m = \{x_1, x_2, x_5, x_6, x_7, x_{13}, x_{14}\}$, while all other vertices can be regarded as functional controllers. To check if what we got in Figure 2C is indeed an MPU, we keep only the variables in the vertex set S_m and their interactions in the evolution equation. Numerical simulation of this reduced set of equations produced an output curve depicted with the thick solid line in Figure 1C, which displays a fast approach to a steady state value that is much larger than the equilibrium value of the full system. It is interesting to note that the saturation value and the relaxation time are very close to those of the first oscillation peak of the full equation. The vertex set S_m constitutes the MPU of the NFκB gene regulation network, and it is the smallest subgraph that generates a quick and large response to the external signal. It can be checked that cutting any link in Figure 2C will totally disrupt the output-producing ability. For example, if the edge (2, 5) (from x_2 to x_5) is cut, the edge (2, 13) has to be cut as well because of the correlation mentioned earlier, and there will be no output signal. The vertices non in S_m act as controllers to bring down the initial pulse to a desired steady value in a larger time scale. Both the short and the long time response in this network bear important biological significance [30].

Biological significance of the MPU and the feedbacks

So far, we have identified the MPU and the feedbacks. Next, we go on to discuss the biological relevance of these "modules" to the operation of NFκB network. In this and several other networks we studied, as an important observation, we find out that the MPU is the core signal production unit which responds quickly to the external cues. In the NFκB network, when a signal such as TNF arrives, IKK$_n$ gets immediately activated into IKK$_a$ while the deactivation of IKK$_a$ is minimized since its constitutive decay rate is small. So, the concentration of IKK$_a$ will rapidly increase until A20 is produced by the feedback loop and starts the catalyzed decay of IKK$_a$. The forward reaction rate is thus maximized transiently and enables cell response to signals with short duration [30]. So, the network has a very sensitive and fast transient response, which is essential for certain signaling pathways [30].

The feedback structures we identified respond at a much larger time scale. Only when the concentration of NFκB reaches a high enough value and induces significant transcriptions in the nucleus, does the negative feedback start to bring down the IKK$_a$ concentration to a steady level which is much lower than the transient peak. The feedback FB$_b$ mainly facilitates the step of clearing NFκB out of the nucleus. FB$_c$ is to restore the concentration of IκBα that has been consumed by the IKK$_a$-catalyzed decay. FB$_d$ is to deactivate IKK$_a$ by A20 to bring down the activation level of the whole network. Thus, our structural decomposition detects forward production unit for quick reaction and feedbacks responsible for long time responses.

Like other feedback signaling from the output [4,38], these loops bring about sensitivity and robustness to the network for fulfilling its basic function [39]. The oscillation observed in Figure 1C is a signature of trading stability for sensitivity [17]. The forward immediate amplification confers easy excitability to the network while together with the delayed feedbacks brings about oscillations. On the other hand, over long time, the reaction rates of all biochemical processes are to some extent influenced by environmental variables such as temperatures, pH values, concentrations of certain ions [40]. To function normally under different conditions, the chemical network should possess structural stability. Here the double feedbacks FB$_c$ and FB$_d$ offer extra structural stability against parameter uncertainty: if the parameter changes incur a temporary increase of the concentration of NFκB$_m$, then both FB$_c$ and FB$_d$ will act to bring it down. Even if one of FB$_c$ or FB$_d$ does not function well, the other one will minimize the change of NFκB concentration. Computation shows that when the rate of the reaction involving either FB$_c$ or FB$_d$ assumes 50% of their normal value, the output signal changes little. However, major changes in the oscillation period, amplitude and the final equilibrium value of the output x_7 are observed when both of the previous changes are made simultaneously. Therefore, these feedbacks provide extra protections for keeping the system stable under parameter fluctuations [22].
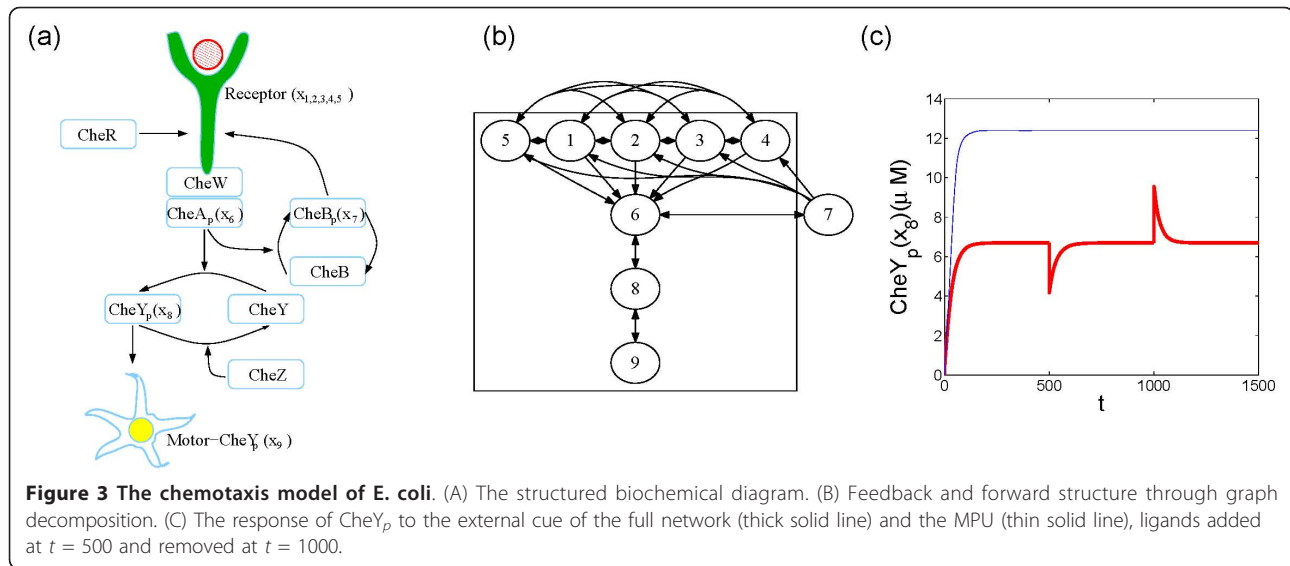
The above procedure of searching for MPU is easily generalized to more complex networks, with possible multiple inputs and outputs which interact with each other. We will study their competition or cooperation all together instead of individually. The critical step lies in our capability of detecting feedback loops. Once the feedback controllers are found, the MPU is obtained by removing all the feedbacks and then all the dynamically inessential nodes. The observed separation of time scales, can, however, leads to further theoretical study using averaging methods or normally hyperbolic invariant manifold concept from dynamical systems. We expect to pursue this in our future studies. In what follows, we analyze the *E. coli* chemotaxis network and several other signaling networks. More examples are available online in the Additional File 1.

Decomposing the E. coli chemotaxis network

Figure 3A displays a chemotaxis model of *E. coli* [41], which enables the *E. coli* cell swimming to food sources and away from hostile environments. The most salient feature of the chemotaxis regulation network is the sensitivity and adaptivity. That is, *E. coli* is able to respond quickly to very weak signals - concentration gradients and under vastly different background concentrations. This special dynamical properties are insured by interesting network topology [4]. As shown in Figure 3A, chemoattractants (indicated by the red ball) bind to and activate the transmembrane receptors ($\{x_1, x_2, x_3, x_4, x_5\}$), which stimulate CheA (x_6) through the adaptor CheW. Activated CheA phosphorylates CheY(x_8), which binds to the flagellar motor (x_9) and increases the frequency of *E. coli* tumbling. The activation of the receptor complex is controlled by its methylation states. Higher methylation states indicate higher probability to be activated. In the model, CheR binds only to the inactive receptors to increase methylation and phosphorylated CheB(x_7) only to the active receptors to decrease methylation.

Figure 3B displays its feedback and forward structure upon application of graph decomposition. The first level consists of the vertex set $\{x_1, x_2, x_3, x_4, x_5\}$ which are different methylation states of the receptor complex. External signals propagate down through x_6, x_8 and finally reaches the flagellar protein x_9.

There is one feedback vertex x_7 (CheB$_p$). The minimal production unit (MPU) is obtained after all the reactions involving x_7 are removed and is contained in the box in Figure 3B.

Figure 3 The chemotaxis model of E. coli. (A) The structured biochemical diagram. (B) Feedback and forward structure through graph decomposition. (C) The response of CheY$_p$ to the external cue of the full network (thick solid line) and the MPU (thin solid line), ligands added at $t = 500$ and removed at $t = 1000$.

With the feedback through CheB$_p$ (x_7), the system has sensitive detection and robust adaptivity as shown with thick solid line in Figure 3C. Starting with zero value, the CheY$_p$ quickly reaches the saturation level. At $t = 500s$, an external stimulus - $10\mu M$ concentration ligand is supplied, which induces a drop of CheY$_p$ concentration followed by an exponential decay back to the saturation value. At $t = 1000s$, the ligand is removed which triggers a jump of CheY$_p$ concentration but regains its stable value exponentially fast. When the feedback is removed, the MPU reaches the stable value after a quick initial rise and stays at the value no matter how the concentration of external ligand changes. The robustness is retained but the adaptivity is lost. So, in this example the feedback is essential for the system's transient response to external stimulus and maintaining the adaptivity. As in the previous example, the production-controller dichotomy structure guarantees the normal functioning of a cell regulation network with both parts playing irreplaceable roles. Here, the forward production reacts quickly accounting for the sensitivity of the network while the controller works in a larger time span to realize the adaptivity.
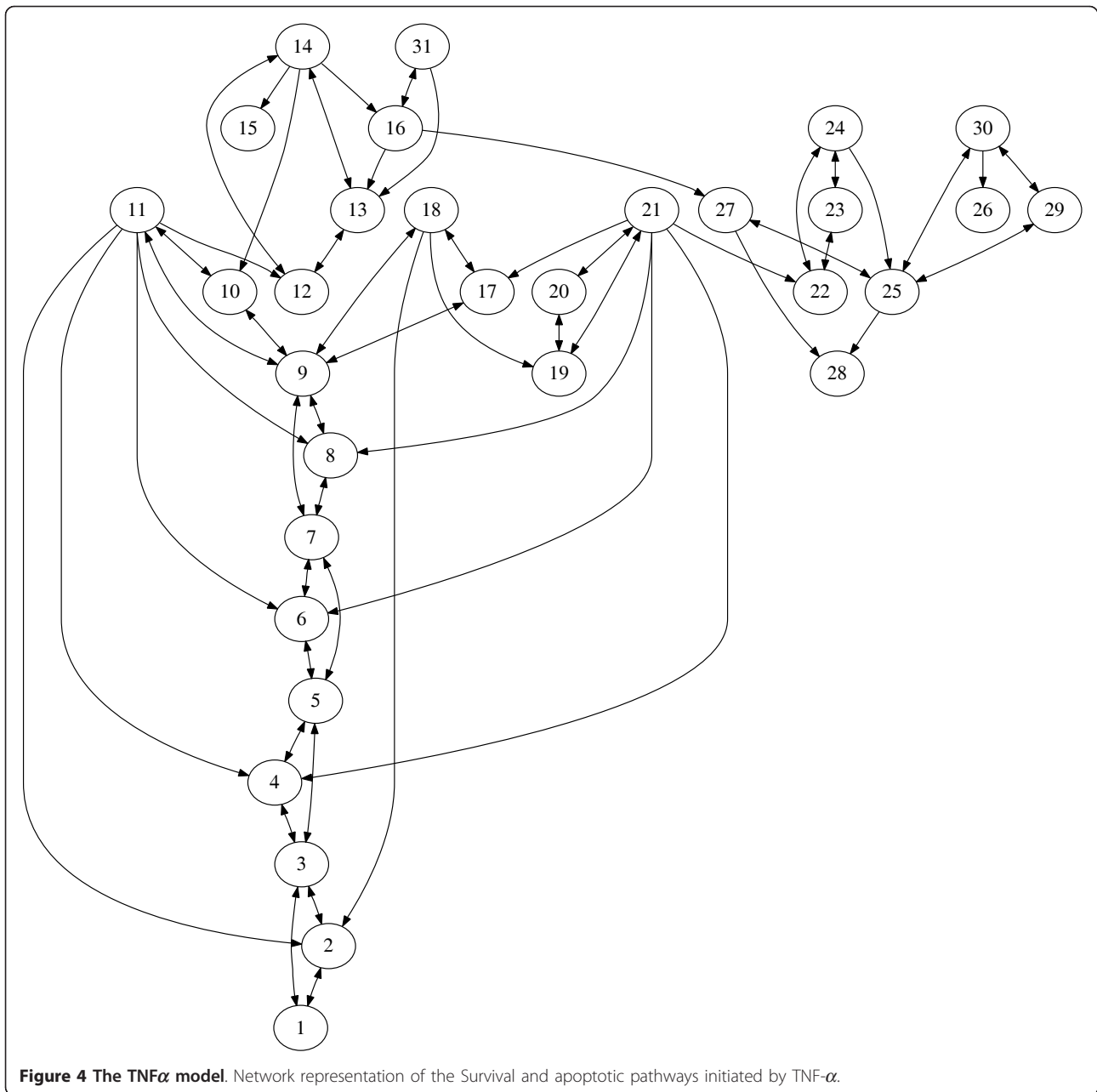
Survival and apoptotic pathways initiated by TNF-α

This model studies the survival and apoptotic pathways initiated by TNF-α that we adopt from [42]. These pathways play decisive roles in cell fate decision in response to inflammation and infection. After an external cue TNF-α binds to its receptor TNFR1 (x_2) (see the table), adaptor proteins TRADD, TRAF2 and RIP-1 are recruited to form an early complex ready for binding and activating other functional proteins. There are two different downstream pathways: the survival pathway

mediated by NF-κB and the apoptotic pathway mediated by caspase. NF-κB is usually sequestered by IκB and is released when IκB degrades. IKK binds to the early complex to form a survival complex and is activated with the dissociation of this complex. The activated IKK is able to induce proteolysis of IκB. The released NF-κB translocates to the nucleus, binds to DNA and leads to the transcription of IAP and IκB. c-IAP inhibits apoptosis by binding to caspase-3* and thus preventing DNA fragmentation. The interaction graph is depicted in Figure 4 and the notation is detailed in Table 1.

Upon application of the graph decomposition routine, we successfully unfold the underlying modular structure of the TNF-α network. The forward production unit is a long cascade involving many different species and reactions. The signal TNF-α (x_1) is processed through the network until DNA fragmentation is induced (x_{26}) as shown Figure 5A. The direct HVD identifies one big SCC enclosed in the two boxes in Figure 5A. Further analysis distinguishes the forward and backward edges. The whole NFκB pathway is now revealed as a feedback module, which controls the level of the c-IAP (x_{27}) and thus Caspase-3* (x_{25}), and maintains the option for survival. It is intriguing that the NFκB module is produced automatically by our decomposition procedure although it has many connections to the rest of the network. The removal of the NFκB module singles out the MPU shown in Figure 5B.

Figure 6 shows the level of DNA fragment (x_{26}) with or without the presence of the NFκB control module. With the feedback module, the rate of the fragmentation of DNA is low (Figure 6A), which may suggest the survival of the cell; without, the DNA cleavage is high (Figure 6B), which could indicate an apoptotic fate of

Figure 4 The TNFα model. Network representation of the Survival and apoptotic pathways initiated by TNF-α.

the cell. So, indeed, here the NFκB modules acts as a controller of the apoptotic pathway. Our decomposition technique accurately captures this information. Again, without the control module, the MPU produces over-abundantly the output signal in a relatively fast way. The long feed-forward edge from x_{16} to x_{27} may acceler-ate the control in this case.

Circadian clock in Drosophila
Circadian clock exists in many different organisms ran-ging from bacteria to human. The regulation pathway

adopted from [43] and displayed in Figure 7 models the *Drosophila* circadian clock which mainly contains two interlocked loops. The notations are explained in Table 2. The TIM and PER protein in the first loop may bind to each other in the cytosol or nucleus, but they enter the nucleus separately. They down-regulate their own expression by inhibiting the transcription factor CLK-CYC. The association of TIM and PER in the cyto-plasm is mediated by FBM and the dissociation is cata-lyzed by SM which is generated by the constitutive entering of PER into the nucleus. In the second loop,

Table 1 The variables in the TNFα model

x_1	TNFα	x_{17}	FADD
x_2	TNFR1	x_{18}	$\langle x_7 \rangle$/RIP1/FADD
x_3	TNFα/TNFR1	x_{19}	TRADD/TRAF2/RIP1/FADD
x_4	TRADD	x_{20}	Caspase8
x_5	TNFα/TNFR1/TRADD	x_{21}	TRADD/TRAF2/RIP1/FADD/ Caspase8
x_6	TRAF2	x_{22}	Caspase8*
x_7	TNFα/TNFR1/TRADD/ TRAF2	x_{23}	Caspase3
x_8	RIP1	x_{24}	Caspase8 * /Caspase3
x_9	μ$\langle x_7 \rangle$/RIP1	x_{25}	Caspase3*
x_{10}	IKK	x_{26}	DNA - frag
x_{11}	$\langle x_7 \rangle$/RIP1/IKK	x_{27}	cIAP
x_{12}	IKK*	x_{28}	Caspase3 * /cIAP
x_{13}	IκB/NFκB	x_{29}	DNA
x_{14}	IκB/NFκB/IKK*	x_{30}	Caspase3 * /DNA
x_{15}	IκBP	x_{31}	IκB
x_{16}	NFκB		

TNFα is one tumor necrosis factor which binds to the receptor TNFR1. TRADD, TRAF2 and RIP1 are adaptor proteins which may form complexes with TNFα. IKK, NFκB, IκB belongs to the NFκB module while FADD, caspase8 and caspase3 are on the apoptotic pathway. c-IAP is an inhibitor of apoptosis protein.

CLK-CYC activates both VRI and PDP expression. VRI represses the expression of CLK while PDP promotes it. Various forms of TIM are also influenced by the sunlight. Nevertheless, even without the coupling to sunlight, the model still produces an oscillation of period 24 hours.

Considering the influence of the external sunlight, we pick x_4 (TIM$_c$) as the input node while x_{21} (CLK·CYC·P$_c$) is selected to be the output node since this complex controls the transcription of TIM and PER. The network graph after the decomposition analysis shown in Figure 8A clearly shows 5 feedbacks. The one through SM (x_{11}) is the positive feedback that accelerates the dissociation of the PER·TIM complex. The other four through x_1, x_2, x_{12}, x_{13} are the important regulators of the concentration of PER, TIM, VRI and PDP through DNA expression and protein translation. The feedbacks through x_{12} and x_{13} interact with each other and control the production of CLK (x_{19}). The MPU is very easily obtained by removing the feedback modules and displayed in Figure 8B, which shows how the (sunlight) signal is picked up at x_4, processed via PER·TIM, CLK·CYC interaction and output at x_{21}.

With all the feedbacks, the *Drosophila* network is able to generate stable oscillations with a period of 24 hours. Indeed, employing the kinetic model in [43] and starting with a somewhat arbitrary condition, the network soon reaches an oscillatory state as shown in Figure 9. Without the feedbacks, all the state variables quickly relax to a steady state, in which concentrations are adjusted

from their initial values quickly in the direction (higher or lower) corresponding to the operating point of the circadian clock. The full network follows this response in the short initial time and then feedbacks take effect to make it oscillate. So, these feedbacks are essential elements for the generation of the circadian cycles. Noticeable in Figure 9 are clearly three distinct time scales: the fastest direct response produced by the MPU, the period of the oscillation and the slowest drift to stable oscillation. The displayed feedbacks in Figure 9 are responsible for the slow adjustment of the motion and the oscillation.
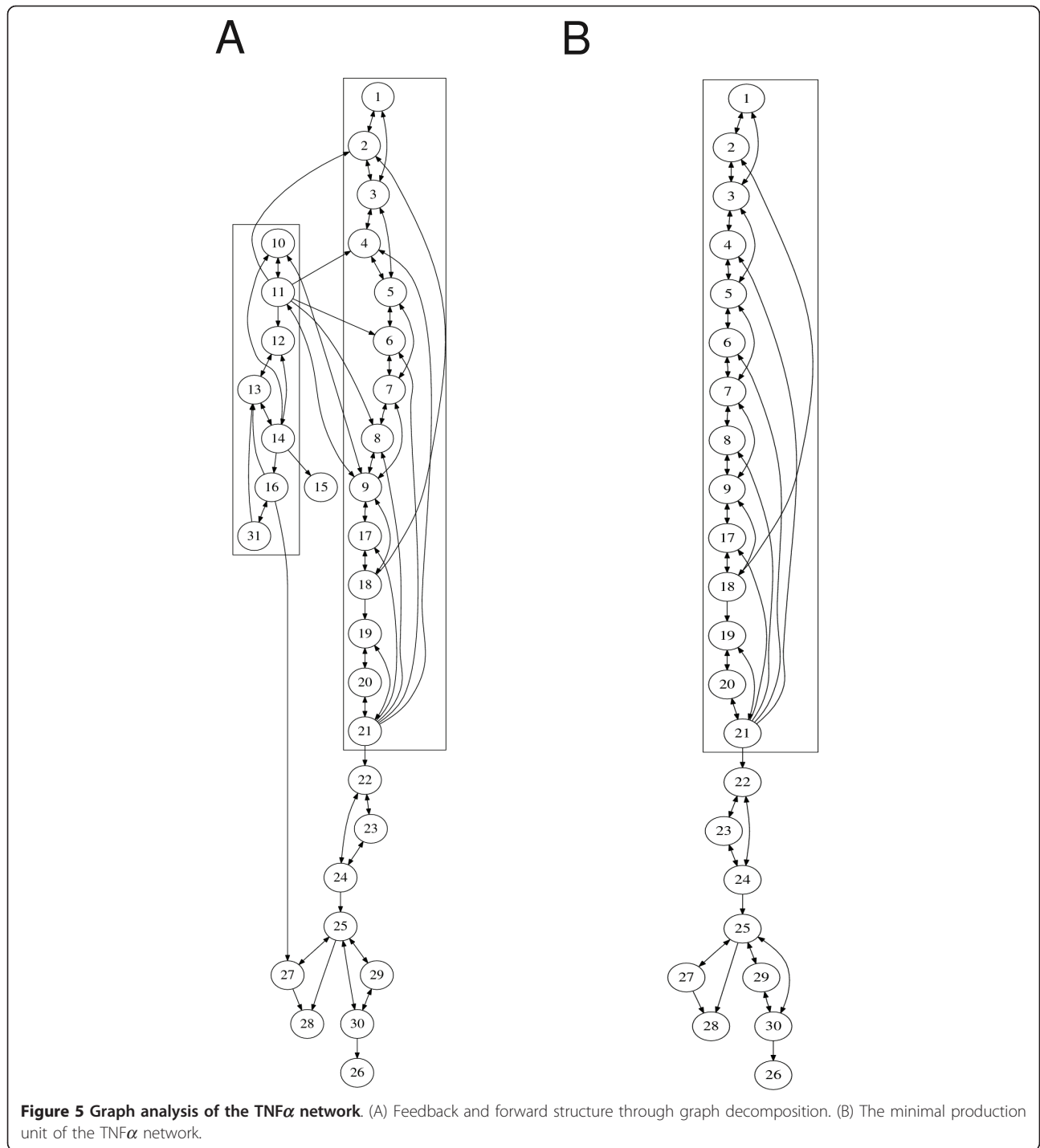
Conclusions

In this paper, we discuss some of the universal aspects of the architecture of biochemical networks that relate to their production and feedback function. We also devise an automatic procedure for identifying the key functional modules of that architecture by applying graph theoretic methods and invoking additional dynamic information. The key ingredients of the architecture are revealed by identifying the forward production unit and the feedback controller. We successfully applied the HVD and the feedback loop searching and selection algorithm and obtained this anatomy in the NFκB regulatory, the *E. coli* chemotaxis network, the TNF-α pathway and the circadian network. In the Additional File 1 we show that similar structures exist in a number of other cell regulatory networks.

The dissection of large networks into functional modules greatly facilitates their analysis. The functional modules can be studied individually with well-designed boundary conditions. The properties of the whole network are deducible by piecing together the modules in an ordered way. Henceforth, our strategy of analysis is characterized by a decomposition and recombination procedure. Current technique can be further extended to the analysis of hierarchical structures at different scales with disparate internal dynamics. In the top-down direction, the network may be broken into functional modules at different scales by the above decomposition technique. From bottom up after the property of each module is conveniently explored, a hierarchy of modules of increasing size may be built until the whole network is covered. From biological evolution point of view, it is likely that this nested structure stems from a simple core and is later wrapped with complex regulation mechanisms during evolution. So, our theory reveals the stable, potentially generic feature of a biochemical network, which can be used to explore either the intricacy in a single structure or interdependencies of a series of systems.

The detection of modular structures provides additional insight into how a regulatory network works and thus gives clear indication of key protein species and key

Figure 5 Graph analysis of the TNFα network. (A) Feedback and forward structure through graph decomposition. (B) The minimal production unit of the TNFα network.

reactions in a cascade, which finds wide applications in the drug design and synthetic biology [44]. The identification of the dominating skeleton subnetwork such as the MPU and key feedbacks in a regulatory pathway also simplifies the determination of reaction rates of *in vivo* biochemical reaction since the distracting unimportant reaction components have been removed from the skeleton structure [45,46]. In all, the production and feedback dichotomy of biological networks shapes cellular signaling [22] and the current graph decomposition technique provides a convenient handle to uncover this important aspect of their architecture.
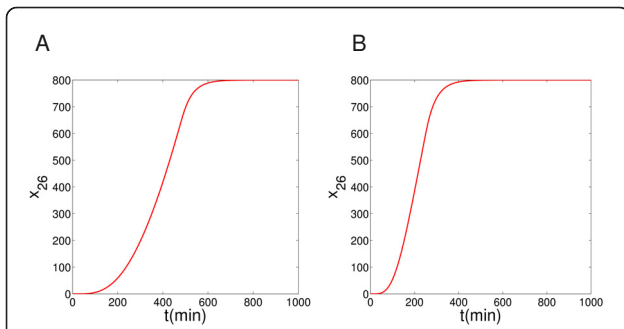
Figure 6 The evolution of the DNA fragment. The evolution of the DNA fragment (x_{26}) (A) with and (B) without the NFκB feedback module.

Methods

Identification of forward and feedback edges

As mentioned previously, here, we present an algorithm to identify the forward and feedback edges with given polarity, by searching and ordering important topological invariants - cycles. First, a cycle search procedure is discussed which produces all the cycle generators for a strongly connected component. Then a selection procedure is discussed which generates a partial order of the vertices and enables the detection of feedbacks in a straightforward way. Before proceeding directly to the
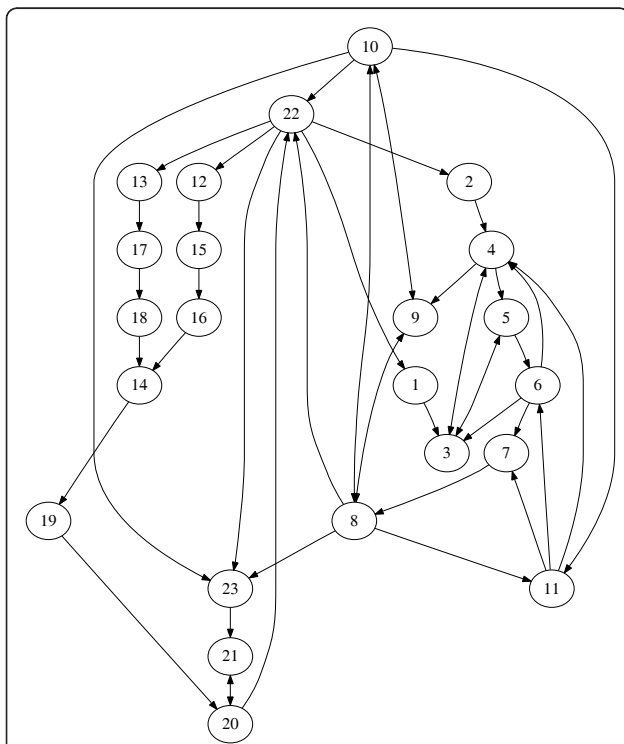
Table 2 The variables of the circadian clock model

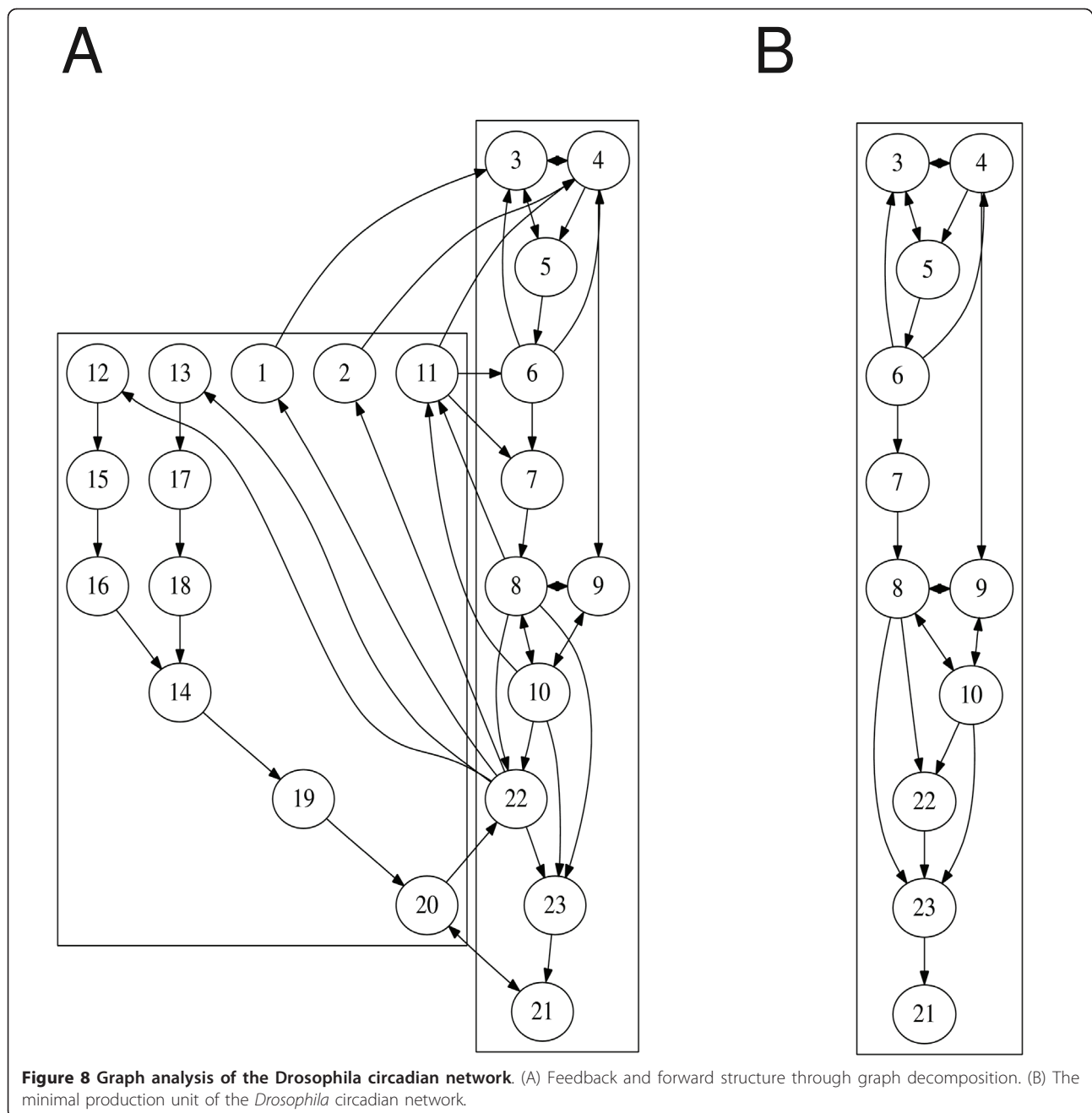| x_1	Per$_m$	x_7	PER · P$_c$	x_{13}	Pdp$_m$	x_{19}	CLK$_c$
x_2	Tim$_m$	x_8	PER · P$_n$	x_{14}	Clk$_m$	x_{20}	CLK · CYC$_c$
x_3	PER$_c$	x_9	TIM$_n$	x_{15}	VRT$_c$	x_{21}	CLK · CYC · P$_c$
x_4	TIM$_c$	x_{10}	PER · TIM$_n$	x_{16}	VRI$_n$	x_{22}	CLK · CYC$_n$
x_5	PER · TIM$_c$	x_{11}	SM$_c$	x_{17}	PDP$_c$	x_{23}	CLK · CYC · P$_n$
x_6	PER · TIM$_f$	x_{12}	VRI$_m$	x_{18}	PDP$_n$		

This *Drosophila* circadian clock model consists of two loops. One contains PER and TIM and the other contains PDP and VRI. They interact through CLK-CYC controlled expression. SM and FBM are two proteins assisting in the first loop.

algorithm part, we state a principle which will be used in our selection procedure.

Principle of minimum feedbacks

Very often, in complex systems, multi-step processes are carried out in a well-ordered sequential way with a small number of feedback controllers modulating the behavior of the system. The cascade structure with minimal number of feedback controls yields balance between robustness and evolvability. It also has the advantage of maximizing operation efficiency and minimizing energy cost. As an analogue, we propose that in order to make optimal use of resources and at the same time maintain necessary stability cells employ a minimum feedback principle: the number of feedback edges should be minimal in a cell regulation network. It seems evolutionarily advantageous to allocate only necessary resources to feedback control. As always happens in biology, there may exist other requirements which weaken this principle. Here, we just stick to this principle which produces reasonable results for all the examples we are looking into so far.

How to find a minimum set of feedback edges is an NP-hard problem in graph theory but there exist approximate algorithms which could do the job relatively fast [47]. It is conceivable that the solution might not be unique. However, extra constraints may help remove some non-uniqueness. From a control theory point of view, the signal transduction network consists of two major components, the information forwarding part and the feedback controller. The forwarding part receives external signal at one end, passing and processing it along different paths, and producing an output at the other end. So, the associated information flow defines a direction on the network. The feedback component modulates the flow by sending downstream signals back to upstream nodes. The identification of these two components is essential for understanding the function of different parts of a network. The problem of searching for the minimal set of feedback arcs has to be consistent with the polarity determined by the information flow. Accordingly, we may restate the problem in an equivalent way: find an ordering of the vertices with the given polarity determined by input and output
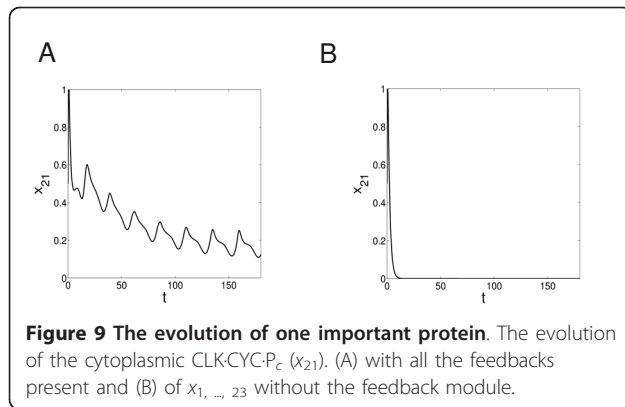

Figure 7 The circadian clock model. Network representation of the Circadian clock in *Drosophila*.

Figure 8 Graph analysis of the Drosophila circadian network. (A) Feedback and forward structure through graph decomposition. (B) The minimal production unit of the *Drosophila* circadian network.

vertices, such that the number of feedback edges is minimized.

Cycle search

For a finite graph, there exist sets of linearly independent cycles, the algebraic combination of which is able to produce all cycles in the graph. Such a set is called the cycle generator set, denoted by Cgen in our paper. It is not unique, but the number of elements in Cgen is fixed for different sets and determined by the graph itself. Therefore, for an SCC, all the edges lie in the

generator set C_{gen}. In the following, we introduce a collapsing scheme to find one C_{gen} of a general graph $\mathscr{G} = \{a_i : i = 1, 2, ..., m | v_i : i = 1, 2, ..., n\}$, where a_i's represent vertices of $\mathscr{G}$, and $v_i = (j, k) = [a_j, a_k]$ represents an edge from the vertex a_j to the vertex a_k. A chain of edges, denoted by $[a_{i_1}, a_{i_2}], \cdots, a_{i_k}]$, represent a path from a_{i_1} to a_{i_2} to ... and finally to a_{i_k}. We use $[a_{i_1}, a_{i_2}]^-, ..., a_{i_k}]$ to represent a cycle of k-nodes. The

Figure 9 The evolution of one important protein. The evolution of the cytoplasmic CLK·CYC·P$_c$ (x_{21}). (A) with all the feedbacks present and (B) of $x_{1, ..., 23}$ without the feedback module.

adjacency matrix of $\mathscr{G}$ is denoted with A. The main idea is to identify shortest cycles and then simplify the graph in an iterative way and a flow graph of the procedure is shown in Figure 10 which is explained below in detail.

(1) Record all the self-loops of $\mathscr{G}$ which are encoded by the nonzero diagonal elements of A. After removing the corresponding edges from $\mathscr{G}$, we obtain a new graph $\mathscr{G}_1$ and a new adjacency matrix A_1.

(2) Search and record a shortest cycle $l_1 = \overline{[a_{i_1}, ..., a_{i_k}]}$ of A_1 for some $k > 1$ by looking for the nonzero diagonal elements of the m-th powers of A.

(3) The induced subgraph $\mathscr{H}_1$ with the vertex set $\{a_1, ..., a_m\}$ and their connections has an adjacency matrix B_1 which is a submatrix of A_1. Each nonzero element (i_p, i_{p+1}) of B_1 can be made to a cycle by connecting $a_{i_{p+1}}$ back to a_{i_p} with part of the cycle l_1, *e.g.*, by the chain of edges $[a_{i_{p+1}}, a_{i_{p+2}}, ..., a_{i_p}]$. Initially, this step is not necessary since besides l_1 there is no extra edge in $\mathscr{H}_1$. However, after the collapse in step 4, there may appear multi-edges between some pair of nodes. For each of those in $\mathscr{H}_1$ but not in l_1, we can identify and record a new cycle.

(4) Collapse all the edges and vertices in the subgraph $\mathscr{H}_1$ into one point P_1, and we obtain the updated graph $\mathscr{G}_2$ for which a new adjacency matrix A_2 is written down. If $\mathscr{G}_2$ only contains P_1, the iteration is terminated. Otherwise, we go back to step 2 and repeat the procedure with the new graph $\mathscr{G}_2$ and the new adjacency matrix A_2. Note that $\mathscr{G}_2$ may not be a simple graph: there could be more than one edge between some pair of vertices. This is the origin of extra edges on a shortest cycle in step 3.
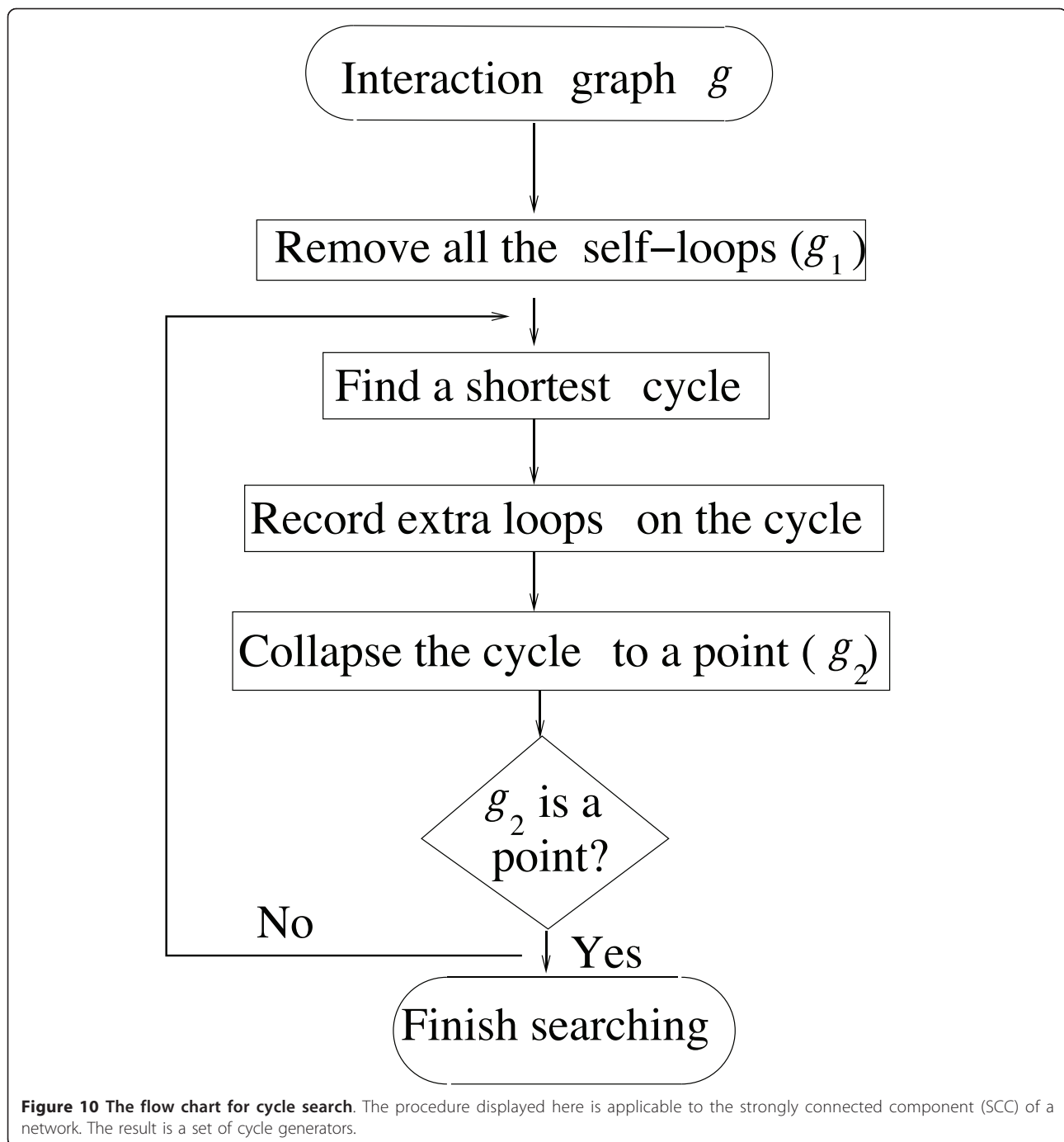
It is easy to show that each cycle of $\mathscr{G}_1$ corresponds to a unique cycle either in $\mathscr{H}_1$ or in $\mathscr{G}_2$. Vice versa, each cycle l in $\mathscr{G}_2$ can be identified with a unique cycle in $\mathscr{G}_1$: if the cycle l runs through P_1, then its incidence vertex and exit vertex in $\mathscr{H}_1$ can be connected by a unique path embedded in the cycle l_1 and thus a unique cycle in $\mathscr{G}_1$ is produced by concatenating this path to the edges contained in l; if the cycle l stands apart from P_1, it directly corresponds to one cycle in $\mathscr{G}_1$. So, after the search is done, finally, we can trace backward all the cycles we have found so far in the original graph $\mathscr{G}$ except the self-loops. In the algorithm just described, not all cycles but a set of linearly independent cycles are recorded, which by definition constitutes a cycle generator set Cgen. The generators derived from the above algorithm are prime in the sense that any proper subset of a generator is not a cycle. Note that the set Cgen may not be unique since the selected cycle in step 2 might not be unique. What consequences this non-uniqueness brings about is an interesting problem that deserves further investigation. However, the important point here is that all the feedback edges appear at least once in Cgen.

For the NF-κB gene regulatory network, we apply the cycle-searching technique and find that the total number of cycle generators are 33 with 15 1-cycles and 8 2-cycles. 10 cycle generators have length greater than 2.
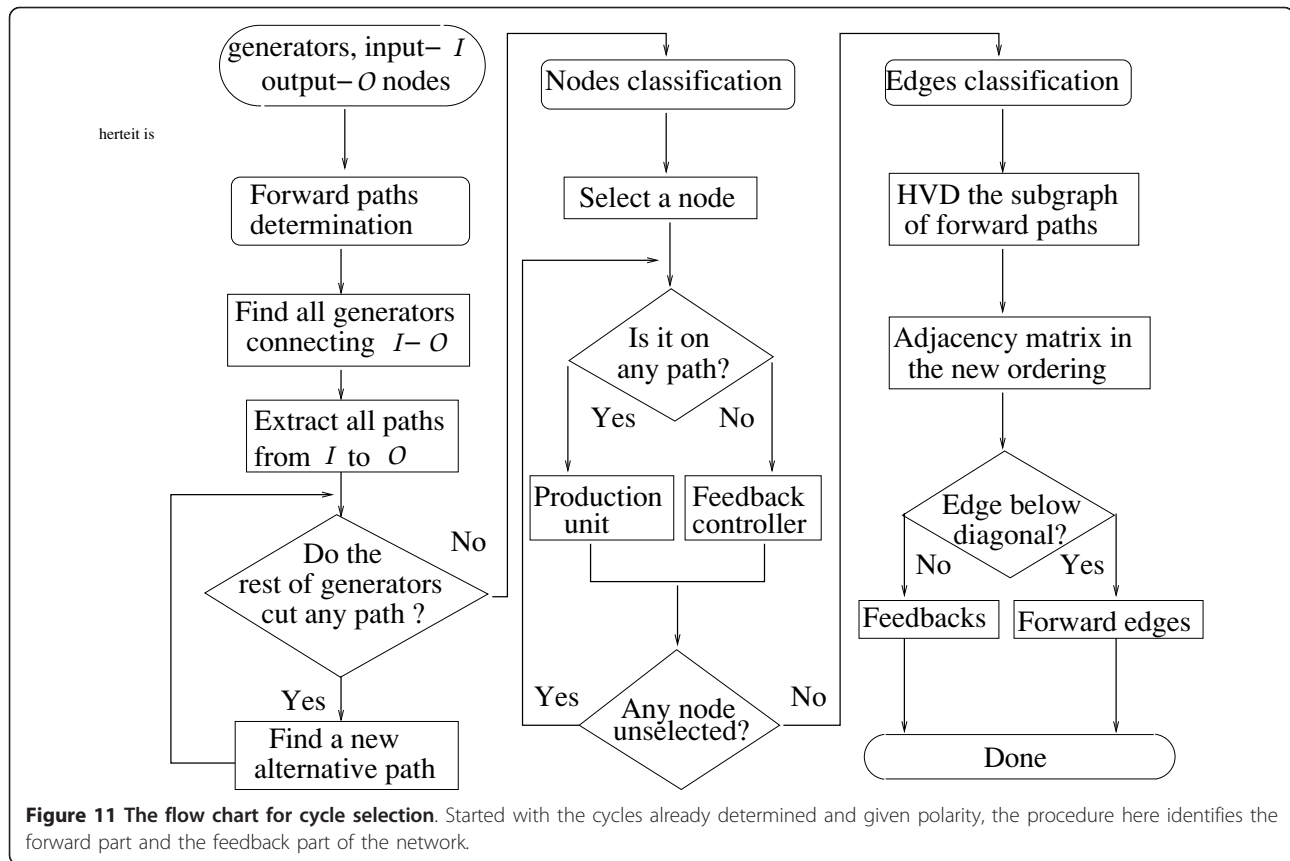
Selection procedure

For a graph with a tree structure, it is always possible to find an order of the vertices, such that only forward edges show up. For instance, the HVD of the graph could generate such an ordering. With cycles present, at least one feedback exists no matter how the vertices are ordered. According the principle of minimal feedbacks, we want to find an ordering under which the number of feedbacks is minimized. That is to say, we set out to extract a minimal set of the edges, the removal of which leads to a cycle-free network. It is not uncommon that by removing one edge quite a few cycles get destroyed. In order to determine the forward and feedback edges in an SCC based on the cycles found in the previous section, we utilize the polarity of the network that determines the flow direction. In a graph possessing polarity, for the input vertex, every out-edge is regarded as a forward edge and every in-edge a feedback. The opposite is true for the output vertex. Note that 1-cycles (self-loops) and 2-cycles are special and need to be treated differently. 1-cycles always attach to individual vertices and are not regarded as feedback loops. 2-cycles are bidirectional edges that are most likely representing the forward and the backward reactions since many of biochemical reactions are reversible. These 2-cycles are important for keeping chemical balance but not usually regarded as

Figure 10 The flow chart for cycle search. The procedure displayed here is applicable to the strongly connected component (SCC) of a network. The result is a set of cycle generators.

feedbacks from a signal transduction point of view. Each 2-cycle contributes exactly one feedback edge and one forward edge, irrespective of the ordering of the vertices. Therefore, they have no impact on the vertex ordering regarding the search of minimal set of feedbacks. Hence, in the selection procedure, we only consider cycles of length greater than or equal to 3 and we call them "long cycles". We take the middle layer of the NFκB network from the HVD result as an example. Here, the input point is x_2 as it receives signals from x_1 and the output point is x_7 as it sends signals to x_{15}. The goal is to find all the forward paths that go from x_2 to x_7 and all the feedback loops. The algorithm is depicted as a flow chart in Figure 11 and a detailed explanation is given below.

Figure 11 The flow chart for cycle selection. Started with the cycles already determined and given polarity, the procedure here identifies the forward part and the feedback part of the network.

(1) With the long cycles and the polarity determined, we first look for cycles connecting x_2 and x_7 and thus extract a set of forward paths that go from x_2 to x_7.

(2) From the remaining long cycles, we search for the ones intersecting an extracted path at two nodes. If more than two intersections are found, we choose the two intersections that are most separated. This choice is to put as many edges as possible to the forward direction and thus to minimize the feedback ones. Using the edges on the cycle as a replacement of the edges in the path that connect the two intersections, an alternative path from x_2 to x_7 is constructed.

(3) We repeat the search until no more alternative paths can be generated from the available long cycles.

(4) Now, it is possible to construct a subgraph $\mathcal{F}$ expanded by the vertices and the edges contained in these forward paths. A node in $\mathcal{F}$ belongs to the production unit and to the feedback controller otherwise. For the middle layer of the NFκB network, the vertex set in $\mathcal{F}$ has been computed as $V_f = \{x_2 , x_5 , x_6 , x_7 , x_{10} , x_{11} , x_{13}\}$ which sit in the forward production unit and are displayed inside the rectangle in Figure 2A. The complementary vertex set consists of $V_b = \{x_4 , x_8 , x_9 , x_{12} , x_{14}\}$ which should be included in the feedback controller.

(5) The HVD is applied to $\mathcal{F}$ to partially order its vertices and edges. We rearrange the order of the vertices in $\mathcal{F}$ according to the partial order. In the new order, an adjacency matrix only has subdiagonal non-zero entries, which represent forward edges. If we restore all edges in the original graph that connect nodes in $\mathcal{F}$, the adjacency matrix may have superdiagonal entries, which are considered as feedback edges. For the NFκB network, the collection of the feedback and the forward edges are clearly seen in the rectangle in Figure 2A. For a complex feedback controller, if needs arise, we may carry out further decomposition with our cycle search and selection algorithm. For the NFκB network, it is not necessary since the feedback controllers are simple line graphs.

Additional material

Additional file 1: Examples of several other cell regulation networks.

Author details

[1]Department of Physics, Tsinghua University, Beijing 100084, China. [2]The Center for Control, Dynamical Systems and Computation, University of California, Santa Barbara, CA 93106, USA. [3]Department of Mechanical Engineering, University of California, Santa Barbara, CA 93106, USA.

Authors' contributions

IM conceived the dichotomy structure of cell regulation network and emphasized the importance of the feedback loops in understanding network function. YL designed the decomposition scheme based on the cycle search and selection and did the analysis on various networks. Both authors read and approved the final manuscript.

References

1. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P: *Molecular Biology of the Cell.* Fourth edition. New York: Garland Science; 2002.
2. You L: **Toward computational systems biology.** *Cell Biochem Biophys* 2004, **40**:167.
3. Endy D, Brent R: **Modelling cellular behavior.** *Nature* 2001, **409**:391.
4. Barkai N, Leibler S: **Robustness in simple biochemical networks.** *Nature* 1997, **387**:913.
5. Lu T, Shen T, Zong C, Hasty J, Wolynes PG: **Statistics of cellular signal transduction as a race to the nucleus by multiple random walkers in compartment/phosphorylation space.** *Proc Natl Acad Sci USA* 2006, **103**:16752-16757.
6. Lan Y, Wolynes P, Papoian GA: **A variational approach to the stochastic aspects of cellular signal transduction.** *J Chem Phys* 2005, **125**:124106.
7. Hasty J, Collins JJ: **Translating the noise.** *Nat Genet* 2002, **31**:13.
8. Swain PS, Elowitz MB, Siggia ED: **Intrinsic and extrinsic contributions to stochasticity in gene expression.** *Proc Natl Acad Sci* 2002, **99(20)**:12795-12800.
9. Kærn M, Elston TC, Blake WJ, Collins JJ: **Stochasticity in gene expression: from theories to phenotypes.** *Nat Rev Genet* 2005, **6**:451.
10. Oda K, Matsuoka Y, Funahashi A, Kitano H: **A comprehensive pathway map of epidermal growth factor receptor signaling.** *Mol Syst Biol* 2005, **2005.0010**:1.
11. Oda K, Kitano H: **A comprehensive map of the toll-like receptor signaling network.** *Mol Syst Biol* 2006, **2006.0015**:1.
12. Barabási AL, Oltvai ZN: **Network biology: understanding the cell's functional organization.** *Nature Rev Gen* 2004, **5**:101.
13. Shen-Orr SS, Milo R, Mangan S, Alon U: **Network motifs in the transcriptional regulation network of *Escherichia coli*.** *Nat Genet* 2002, **31**:64.
14. Goldbeter A: **A minimal cascade model for the mitotic oscillator involving cyclin and cdc2 kinase.** *Proc Natl Acad Sci USA* 1991, **88**:9107-9111.
15. Kaern M, Blake WJ, Collins JJ: **The engineering of gene regulatory networks.** *Annu Rev Biomed Eng* 2003, **5**:179-206.
16. Freeman M, Gurdon JB: **Regulatory principles of developmental signaling.** *Annu Rev Cell Devel Biol* 2002, **18**:515-539.
17. Csete ME, Doyle JC: **Reverse engineering of biological complexity.** *Science* 2002, **295**:1664.
18. Ravasz E, Barabási AL: **Hierarchical organization in complex networks.** *Phys Rev E* 2003, **67(2)**:026112.
19. Watts DJ, Strogatz SH: **Collective dynamics of 'small-world' dynamics.** *Nature* 1998, **393**:440.
20. Mason O, Verwoerd M: **Graph theory and networks in biology.** *IET Syst Biol* 2007, **1(2)**:89-119.
21. Kuo PD, Banzhaf W, Leier A: **Network topology and the evolution of dynamics in an artificial genetic regulatory network model created by whole genome duplication and divergence.** *BioSystems* 2006, **85**:177-200.
22. Brandman O, Meyer T: **Feedback loops shape cellular signals in space and time.** *Science* 2008, **322**:390.
23. Lewis J: **From signals to patterns: space, time and mathematics in developmental biology.** *Science* 2008, **322**:399.
24. Clauset A, Moore C, Newman MEJ: **Hierarchical structure and the prediction of missing links in networks.** *Nature* 2008, **06830**:98.
25. Alon U: **Network motifs: theory and experimental approaches.** *Nature* 2007, **8**:450.
26. Hoffmann A, Leung TH, Baltimore D: **Genetic analysis of NF-κB transcription factors defines functional specificities.** *EMBO J* 2003, **22(20)**:5530.
27. Rangamani P, Sirovich L: **Survival and apoptotic pathways initiated by TNF-α: modeling and predictions.** *Biotech Bioengr* 2006, **97**:1216.
28. Kuczenski RS, Hong KC, García-Ojalvo J, Lee KH: **PERIOD-TIMELESS interval timer may require an additional feedback loop.** *Plos Comput Biol* 2004, **3**:1468.
29. Lipniacki T, Paszek P, Brasier AR, Luxon B, Kimmel M: **Mathematical model of NF-κB regulatory module.** *Biophys J* 2004, **228**:195-215.
30. Hoffmann A, Levchenko A, Scott ML, Baltimore D: **The I$_κ$B-NRκB signaling module: temporal control and selective gene activation.** *Science* 2002, **298**:1241.
31. Scheidereit C: **Iκ B kinase complexes: gateways to NF-κB activation and transcription.** *Oncogene* 2006, **25**:6685.
32. Fujarewicz K, Kimmel M, Swierniak A: **On fitting of mathematical models of cell signaling pathways using adjoint systems.** *Math Biosci Engr* 2005, **2(3)**:527-534.
33. Mezić I: *Coupled nonlinear dynamical systems: asymptotic behavior and uncertainty propagation* 2004, [43rd IEEE Conference on Decision and Control].
34. Craciun G, Feinberg M: **Multiple equilibria in complex chemical reaction networks: I. the injectivity property.** *SIAM J Appl Math* 2005, **65(5)**:1526.
35. Craciun G, Feinberg M: **Multiple equilibria in complex chemical reaction networks: II. the species reaction graph.** *SIAM J Appl Math* 2006, **66(4)**:1321.
36. Tarjan R: **Depth first search and linear graph algorithms.** *SIAM J Comput* 1972, **1**:146-160.
37. Kitano H, (Ed): *Foundations of Systems Biology* Cambridge: The MIT Press; 2001.
38. Barkai N, Leibler S: **Circadian clocks limited by noise.** *Nature* 2000, **403**:267.
39. Cheong R, Bergmann A, Werner SL, Regal J, Hoffmann A, Levchenko A: **Transient IκB kinase activity mediates temporal NFκB dynamics in response to a wide range of Tumor Necrosis Factor-α doses.** *J Biol chem* 2006, **281(5)**:2945.
40. Ihekwaba AEC, Wilkinson SJ, Waithe D, Broomhead DS, Li P, Grimley RL, Benson N: **Bridging the gap between *in silico* and cell-based analysis of the nuclear factor-κB signaling pathway by *in vitro* studies of IKK2.** *FEBS J* 2007, **274**:1678.
41. Rao CV, Kirby JR, Arkin AP: **Design and diversity in bacterial chemotaxix: a comparative study in *Escherichia coli* and *Bacillus subtilis*.** *PLoS Biol* 2004, **2**:0239.
42. Rangamani P, Sirovich L: **Survival and apoptotic pathways initiated by TNF-α: modeling and predictions.** *Biotech Bioengr* 2006, **97**:1216.
43. Kuczenski RS, Hong KC, García-Ojalvo J, Lee KH: **PERIOD-TIMELESS interval timer may require an additional feedback loop.** *PLoS Comput Biol* 2007, **3**:1468.
44. Arkin AP: **Synthetic cell biology.** *Curr Opin Biotech* 2001, **12**:638-644.
45. Golub G, van Loan C: *Matrix Computations* Baltimore, Maryland: Johns Hopkins University Press; 1996.
46. Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP: **Universally sloppy parameter sensitivities in systems biology models.** *PLoS Comput Biol* 2007, **3(10)**:e189.
47. Walther H: *Ten applications of graph theory* Dordrecht: D. Reidel Publishing Company; 1984.