Theory on the Dynamics of Feedforward Loops in the Transcription Factor Networks

Rajamanickam Murugan*

Department of Biotechnology, Indian Institute of Technology Madras, Chennai, India

Abstract

Feedforward loops (FFLs) consist of three genes which code for three different transcription factors A, B and C where B regulates C and A regulates both B and C. We develop a detailed model to describe the dynamical behavior of various types of coherent and incoherent FFLs in the transcription factor networks. We consider the deterministic and stochastic dynamics of both promoter-states and synthesis and degradation of mRNAs of various genes associated with FFL motifs. Detailed analysis shows that the response times of FFLs strongly dependent on the ratios \( \frac{\gamma_{pd}}{\gamma_{ph}} \) where \( h = a, b, c \) corresponding to genes A, B and C between the lifetimes of mRNAs \( \frac{1}{\gamma_{pd}} \) of genes A, B and C and the protein of C \( \frac{1}{\gamma_{ph}} \). Under strong binding conditions we can categorize all the possible types of FFLs into groups I, II and III based on the dependence of the response times of FFLs on \( \gamma_{pd} \). Group I that includes C1 and I1 type FFLs seem to be less sensitive to the changes in \( \gamma_{pd} \). The coherent C1 type seems to be more robust against changes in other system parameters. We argue that this could be one of the reasons for the abundant nature of C1 type coherent FFLs.

Introduction

Transcription factors (TFs) regulate the quantitative levels of many proteins inside a cell [1–4]. TF networks consist of several fundamental building blocks such as auto regulatory loops, flip-flops, feedback loops, single input modules, cascades, feed-forward loops (FFL) and dense overlapping regulons [5–7]. Positive auto regulatory loops play critical roles in the maintenance of cellular memory [3] and reprogramming whereas a negative auto regulatory loop seems to speed up the response time against an external stimulus [8–9]. The response time of a gene/motif is the amount of time that is required to achieve half of the steady-state concentration of resultant protein product which is often referred to as rise-time [3–4].

Feedforward loops consist of three different genes namely A, B and C which code for three different TFs. Here the protein of gene A regulates the transcription of both B and C whereas both the proteins of genes A and B regulate the transcription of C (Fig. 1). As shown in Figure 1, totally there are three such regulatory connections in a FFL network motif and eight such regulatory combinations viz. [PPP, PNP, NPN, PNN, PPP, NNP, and NNP]. The first one in a combination “FGH” denotes the type of regulation of the transcription of gene B by the protein of gene A, the second one stands for the type of regulation of C by B and the third one denotes the type of regulation of C by TF protein A. Here the type of regulation can be either positive or negative where P denotes positive and N denotes negative. Here PPP is classified as a coherent type FFL whereas PNP is classified as incoherent type. A FFL motif is said to be a coherent type if the direct effect of the general transcription factor (A) on the effector operons (C) has the same sign (positive or negative) as its net indirect effect through the specific transcription factor (B). Here PPP, NPN, PNN and NNP (termed as C1, C2, C3 and C4) are coherent types and PNP, NNN, PPN and NPP (termed as I1, I2, I3 and I4) are incoherent types [10–11]. Here A is the general transcription factor that directly regulates the effector operon of gene C and also indirectly regulates gene C through B. For example, in PNN coherent type, gene A positively regulates B which in turn negatively regulates C and therefore the net effect of regulation of C by TF gene A indirectly through B is negative. Since A directly regulates C via negative mode, PNN is called as a coherent type.

FFLs perform several important cellular tasks in various biological systems [11–15]. It has been shown that PPP type FFL can act as a sign sensitive delay [14]. For example when A and B regulate C via an AND type logic, then PPP type FFL shows a delay in the expression of gene C following induction of A by an external signal, but no delay following deactivation of A [14]. It seems that coherent types constitute ~85% of the naturally occurring FFL motifs. Magnan and Alon [10] have comprehensively studied the dynamical and kinetic behavior of various types of FFLs under the condition that the promoters of both genes A and B were triggered by external stimuli/signals. Upon analyzing literature-based databases of experimentally verified direct transcription interactions for *E. coli* [6] and *S. cerevisiae* [7], they have discovered that PPP (C1) and PNP (I1) type FFLs are more abundant in nature than others [10]. Although all the FFLs are biologically feasible [10], it is still not clear why these two types were preferably selected by nature against other types in these organisms.

Most of the earlier theoretical and experimental studies on FFLs assumed a quasi-equilibrium condition for the binding-unbinding
dynamics of regulatory TF proteins at various promoters [10] and a steady-state condition for the dynamics of synthesis and degradation of mRNAs. These assumptions are valid [9] only when the timescales associated with the synthesis and degradation of TF proteins are much slower (several orders of magnitude) than the timescales associated with the binding-unbinding of regulatory TFs at the respective promoters and synthesis and degradation of mRNAs. Recently the role of mRNA stability in tuning the kinetics of gene induction has been studied in detail by Elkon et al [16]. It seems that the rapidity of induction negatively correlates with the stability of mRNAs. Further, the dynamics of mRNAs can be approximated to be in a steady state only when the ratio \( w = \frac{\gamma_p}{\gamma_m} \) (= lifetime of mRNA/lifetime of protein) is closer to zero which is not true for most of the protein coding genes [9]. Here \( \gamma_p \) and \( \gamma_m \) are the decay rate constants associated with the respective protein and mRNA. In prokaryotic systems \( w \sim 0.1 \) and in eukaryotic systems such as yeast \( w \) seems to vary [17–19] approximately from 0.1 to 1 with a median of \( \sim 0.3 \). The response times associated with various TFs in a given network are strongly dependent on \( w \) of the respective genes and the rate of protein production will be in turn dependent on the response time [9]. TF regulatory networks whose response times are close to or lesser than the generation time of the cell and also less sensitive to the variation in the values of \( w \) corresponding to the component genes are more desirable. In this paper using a combination of theoretical and simulation tools we will (a) formulate a detailed model of various types of FFLs that includes the binding-unbinding dynamics of regulatory TFs at various promoters and synthesis and degradation of mRNAs, (b) investigate the effect of variation in \( w \) and other system parameters on the response times and overall dynamics of different type of FFLs using the detailed model and (c) explain why some of the FFLs are more abundant in nature than others.

**Results**

**Theoretical Formulation**

Feedforward loops consist of three genes coding for transcription factors (TF) A, B and C (Fig. 1). The corresponding

![Figure 1. Various types of feedforward loops (FFLs) considered. FFLs consist of three genes which code for three different transcription factors A, B and C where B regulates C and A regulates both B and C. There are three regulatory connections in FFLs. Since each of these regulatory connections can be either positive or negative, totally there are eight different FFLs. We use three letter codes as “FGH” where ‘F’ denotes the type of regulation of B by A and ‘G’ denotes the type of regulation of C by B and ‘H’ denotes the type of regulation of C by A. Here PPP, NPN, PNN and NNP (termed as C1, C2, C3 and C4) are coherent type and PNP, NNN, PPN and NPP (termed as I1, I2, I3 and I4) are incoherent type. When dimer of A and B regulate C through AND-logic, there are four possible FFLs as P-P, P-N, N-N and N-P. doi:10.1371/journal.pone.0041027.g001](doi:10.1371/journal.pone.0041027.g001)
concentrations of mRNAs are (m_c, m_a, m_b) and the concentrations of the protein products are p_a, p_b and p_c are all measured in mol/lit [M]. Corresponding steady-state concentrations are denoted as (m_{c}, m_{a}, m_{b}, p_{a}, p_{b}, and p_{c}). In the absence of any regulation or when the promoters are turned-on completely, these steady state values will be (m_{c} = k_{m_{c}}/\gamma_{mh}, p_{a} = k_{m_{a}p_{a}}/\gamma_{ah}p_b\gamma_{bc}, p_{b} = k_{m_{b}p_{b}}/\gamma_{bh}p_c\gamma_{cb}).

Here subscripts h = a, b, c. Here the transcription rate of TF gene 'h' is denoted as (s^{-1}) and the respective translation rate is (s^{-1}). The decay rate constant corresponding to the protein product is (s^{-1}). Here subscripts h, g = a, b, c respectively denote gene A, B and C.

The gene associated with the transcription factor A is controlled/triggered by external signal which may be an arbitrary time dependent pulse function (we denote this as \gamma(t)) or an exponentially decaying one. There is a cis-acting element associated with the promoter of gene B where TF protein of A can bind and hence up/down regulate the expression of B via distal action that is mediated by either tracking or looping modes [20]. There are cis-acting elements associated with promoter of C where the protein products of both genes A and B or the dimer of A-B can bind and hence can up/down regulate C respectively in a “OR” or “AND” logic mode. In an A-OR-B mode the presence of either protein A or B is enough to up/down regulate the promoter of gene C. The protein products of TF gene A and B can also up/down regulate C in a “AND” type logic when the dimer of protein products A-B binds with the promoter of gene C. In this case the presence of both A and B is essential for up/down regulation of gene C. There are eight numbers of regulatory combinations with A-OR-B type logic and four different combinations are possible with A-AND-B type logic (Fig. 1).

The fraction occupancy of promoter of TF gene 'h' by the respective regulatory TF protein 'g' is denoted as (X_{hg} \in (0,1)) which is the ratio (x_{hg}/d_{hg}) where (x_{hg}) is the cellular concentration of the promoter of gene 'h' that is bound with the protein of TF gene 'g' and (d_{hg}) is the total concentration of the promoter of gene 'h' inside the cellular volume. There are at least two types of inter-molecular interactions [22] involved in the binding of TF proteins at the corresponding cis-regulatory sequences namely (a) a weak non-specific electrostatic interactions between the negatively charged backbone of DNA and the positively charged side chains of the aminoacids which are present at the DNA binding domains of TF proteins and (b) the specific hydrogen bonding interactions at the site-specifically bound DNA protein interface. The strength of electrostatic interactions will be modulated by the presence of water molecules at the DNA protein interface. In such well hydrated conditions at the DNA-protein interface, the net electrostatic interactions can be either attractive or repulsive owing to the presence of multiple electrical double layers around each charged group [22]. The free energy barrier associated with the fluctuating dynamics of TF proteins within this electrostatic field is comparable with that of the thermal free energy that in turn helps the TF proteins to freely slide along the DNA within this electrostatic capturing domain without physical dissociation [23].

Dissociation or unbinding (X_{hg} = 0) of TF protein from the cis-acting site happens when all the specific hydrogen bonds are broken and the TF protein completely escapes from this electrostatic force field or capturing domain. When the TF protein is well within the electrostatic field then depending on the net inter-molecular interactions at the DNA-protein interface we find that (X_{hg} \in (0,1)). This is reasonable since there may be a situation where the site-specific hydrogen bonding network present at the interface of cis-regulatory DNA sequence and TF is broken due to thermal induced fluctuations but the TF protein is still present there within the electrostatic force field (partially bound condition). This partially bound promoter state is common in eukaryotic systems where a DNA loop connects the cis-acting module and promoter and holds the transcription initiation components so that they are nearby each other in three dimensional space within the electrostatic capturing domain. Further the binding-unbinding dynamics of TF protein 'g' at the promoter of TF gene 'h' will be observed as a continuous process in the timescale of the synthesis and decay of the corresponding mRNA and protein of TF gene 'h'.

This means that at the timescales of synthesis and decay of mRNAs and proteins, the thermally driven local fluctuations in the occupancy of the promoters by the TF proteins well within the electrostatic capturing domain will be averaged out. Under such conditions the averaged promoter state occupancy will be equal to the thermodynamic probability of finding the promoter to be occupied by the regulatory TF protein. Upon considering all these facts, one can conclude that it is appropriate to use a continuous type probability variable (such as (X_{hg})) to denote the promoter state occupancy to account for the promoter of gene 'h' that is partially bound with the TF protein 'g' rather than a discrete Boolean type variable as described earlier [9,17–10]. The bimolecular collision rates associated with the binding of TF protein 'g' with the promoter of gene 'h' is denoted as (k_{gh} (M^{-1} s^{-1})) and the corresponding off-rates of these bimolecular site-specific DNA-protein complexes are denoted as (k_{gh} (s^{-1})). (K_{m_{hg}} = k_{gh}/k_{gh} (M)) is the overall dissociation constant corresponding to the site-specific binding of TFs at their respective cognate sites. We have summarized the parameters of our detailed model in Tables 1 and 2. Since there are several system parameters, exploration of the entire parametric hyperspace will be a complicated one. To simplify the calculations further we introduce the following scaling scheme to project the dynamical variables onto a dimensionless space.

\[
\tau = \gamma_{pa} \tau; \quad M_k = m_k/m_{ks}; \quad m_{ks} = k_{m_k}/\gamma_{mk}; \quad P_k = p_k/p_{ks};p_{ks} = k_{m_k}k_{pk}/\gamma_{mk}\gamma_{pk}; \quad k = a,b,c
\]

Here we measure the real time \(t\) in terms of numbers of lifetimes (1/\gamma_{pa}) of the protein product of gene C. When protein C is stable over several generations of the cell, then its rise-time that is required to achieve half of the steady state value will be equal to the generation time of the cell upon considering the dilution owing to the doubling of cell volume along the process of cell division [8–9]. In such conditions one can also transform the dimensionless \(\tau\) in terms of numbers of generation times of the cell by dividing as \(t/\tau\). With these definitions we can write the deterministic differential equations associated with temporal evolution of \(M_k, P_k\) and \(X_{hg}\) where \(h = h, g\) and \(g = a, b, c\) of FFLs with A-OR/AND-B type regulatory logic imposed on the transcription of gene C as follows.

**TF A:**

\[
\begin{align*}
\frac{d}{dt}M_a &= \gamma_{pa} - \gamma_{mh}M_a \\
\rho \frac{d}{dt}P_a &= \gamma_{mh}P_a - \gamma_{ah}P_a(X_{bc}) - \gamma_{bh}P_a(X_{cb}) - \gamma_{mh}P_a \quad (1)
\end{align*}
\]

The first one in Eqs 1 describes the dynamics of mRNA associated with the TF gene A whereas the second one describes the dynamics of protein synthesis, decay and binding-unbinding of protein A with the promoters of TF genes B and C. The
We can set on/off the expression of gene A. For a constitutive expression of TF gene A we can set

\[ \psi_{\text{xt}}(\tau) = \frac{\gamma_{\text{xt}}}{\gamma_{\text{ma}}} \]

Here \( \psi_{\text{xt}}(\tau) \in (0,1) \) is a time dependent external signal that can turn on/off the expression of A. To investigate the properties of the response-times of various types of FFLs we can set \( \psi(\tau) \) as rectangular pulses/dips with predefined widths. Here \( \psi(\tau) \) is the Heaviside step function such that \( \psi(\tau) = 0 \) when \( \tau < 0 \) and \( \psi(\tau) = 1 \) when \( \tau > 0 \). Rectangular pulses of signals at a given time point can be constructed with a combination of step functions. To introduce a rectangular pulse at the scaled time \( \tau = \tau_p \) for a width of \( \theta_i \) where \( i = 1, 2 \ldots n \) we need to set

\[ \psi(\tau) = \frac{\gamma_{\text{xt}}}{\gamma_{\text{ma}}} \]
The first one in Eqs 2 describes the binding-unbinding dynamics of TF protein A at the promoter of TF gene B. Second and third equations describe respectively the dynamics of synthesis and degradation of mRNA and protein products associated with TF gene B. The dimensionless perturbation parameters in Eqs 2 are defined as follows.

\[
\begin{align*}
\nu_{ab} &= \gamma_{pc}/k_{fabc}p_{ab};
\nu_{cb} &= \gamma_{pc}/k_{fabc}p_{bc};
\mu_{bc} &= K_{bc}/p_{bc};
\mu_{ac} &= K_{ac}/p_{ac};
\end{align*}
\]

The function \( \Omega_{+}(0, 1) \) will vary depending on the type of regulation. For a positive regulation of the promoter of TF gene B by TF protein of A, we find \( \Omega_{+} = X_{ba} \) and for a negative type regulation we find \( \Omega_{-} = 1 - X_{ba} \).

**TF C:**

\[
\begin{align*}
\nu_{ca} &= \gamma_{pc}/k_{fabc}p_{ca};
\nu_{cb} &= \gamma_{pc}/k_{fabc}p_{cb};
\nu_{bc} &= \gamma_{pc}/k_{fabc}p_{bc};
\end{align*}
\]

The first set of equations in Eqs 3 describe respectively the binding-unbinding dynamics of TF proteins A and B at the corresponding cis-regulatory elements associated with the promoter of gene C. Second and third equations describe the dynamics of synthesis and degradation of mRNA and protein products of C. The dimensionless perturbation parameters in Eqs 3 are defined as follows.

\[
\begin{align*}
\nu_{ca} &= \gamma_{mc}/k_{fabc}p_{ca};
\nu_{cb} &= \gamma_{mc}/k_{fabc}p_{cb};
\nu_{bc} &= \gamma_{mc}/k_{fabc}p_{bc};
\end{align*}
\]

The function \( \zeta_{a+b \pm} = (X_{ca}, X_{cb}) = \zeta_{a-b \pm} \) varies depending on the type of regulation. Here the total fraction of promoter of C occupied by the proteins of either A or B is \( X_{c} = X_{ca} + X_{cb} \). There are four different possibilities.

\[
\begin{align*}
\zeta_{a+b \pm} = \begin{cases} X_{ca} + X_{cb} & \text{if } a+b = + \text{ and } a-b = \pm \text{ or } \text{vice versa} \\
X_{ca} - X_{cb} & \text{if } a+b = - \text{ and } a-b = \pm \text{ or } \text{vice versa} \\
(1 + X_{c})/2 & \text{if } a+b = + \text{ and } a-b = + \\
(1 - X_{c})/2 & \text{if } a+b = - \text{ and } a-b = - \end{cases}
\end{align*}
\]

Here the subscript “a+b” indicates the case where both the TF proteins A and B positively regulate C and other combinations are defined in the similar way. In Eqs 4 we have defined \( X_{c} = X_{cb} - X_{ca} \) and depending on the type of regulation imposed on the promoter of gene C we find the following limiting conditions of the promoter-state occupancy of TF gene C.

\[
\begin{align*}
\lim_{p_{ab} \to 0} \zeta_{a+b} &= 1; \lim_{p_{ab} \to \infty} \zeta_{a-b} = 0; \\
\lim_{p_{ab} \to 0} \zeta_{a-b} - \zeta_{a+b} &= 1/2
\end{align*}
\]

Upon combining Eqs 1–4 we find that there are eight possible numbers of regulatory combinations as (PPP, NNN, PNP, NPN, NNP, PPN, NPP and PNN) in case of A–OR–B type FFL. Here PPP, PNP, NNN and NNP are coherent type FFLs (corresponding standard notations are namely C1, C2, C3, and C4) whereas PNP, NNN, PPN and NPP (corresponding standard notations are namely I1, I2, I3 and I4) are incoherent types FFLs. There will be an additional step corresponding to the dimerization of TF proteins A and B in case of FFLs with A–AND–B logic type regulation. The dynamics of protein-protein dimerization and dissociation can be described by the following differential equation.

\[
\begin{align*}
\varepsilon_{ab} dY_{ab}/dt &= P_{a}P_{b} - L_{ab}Y_{ab} - \sigma_{y}(Y_{ab}(1 - X_{cy}) - \mu_{yc}X_{cy})
\end{align*}
\]

Here \( Y_{ab} \) is the cellular concentration (M) of the dimer of the protein products of TF genes A and B, \( L_{ab} \) and \( \lambda_{ab} \) are the corresponding forward bimolecular (M–1s–1) and reverse unimolecular (s–1) rate constants associated respectively with the dimerization and dissociation reactions. To be consistent with Eqs 6 the equations associated with the dynamics of synthesis and degradation of Pa and Pb will be modified as follows.

\[
\begin{align*}
\rho_{a} dP_{a}/dt &= M_{a} - P_{a} - \sigma_{a}(P_{a}(1 - X_{ba}) - \mu_{ab}X_{ba})
\end{align*}
\]

Further when the dimer A–B binds with the promoter of gene C and hence up/down regulate then the related rate equations corresponding to the expression of gene C will be modified as follows.

\[
\begin{align*}
\varepsilon_{ab} dX_{cy}/dt &= Y_{ab}(1 - X_{cy}) - \mu_{yc}X_{cy}
\nu_{cy} dM_{c}/dt &= \Psi_{y \pm}(X_{cy}) - M_{c}
\end{align*}
\]

The function \( \Psi_{y \pm} \) is the cellular concentration (M) of the dimer of the protein products of TF genes A and B, \( L_{ab} \) and \( \lambda_{ab} \) are the corresponding forward bimolecular (M–1s–1) and reverse unimolecular (s–1) rate constants associated respectively with the dimerization and dissociation reactions. To be consistent with Eqs 6 the equations associated with the dynamics of synthesis and degradation of Pa and Pb will be modified as follows.
\[ \rho_a dP_a/d\tau = \chi(t) - P_a; \quad \rho_b dP_b/d\tau = \Omega_+ (X_{bas}) - P_b; \]
\[ dP_c/d\tau = \zeta_{ab+bc} (X_{cas}X_{chs}) - P_c \]

Eqs 9 can be obtained by setting \( w_h = 0 \) for all \( k = (a, b, c) \) and \( v_c = v_a = v_b = 0 \) in Eqs 1–3. The steady state values of the protein products \( P_h \), where \( h = a, b, c \) will vary depending on the type of regulation and protein-protein interactions between \( A \) and \( B \). Here we have defined the steady state values of \( X_{ab} \) in the coupled Eqs 9 as follows.

\[ X_{bas} = P_a / (\mu_ab + P_a); \quad X_{cas} = \mu_bP_a / (\mu_abP_b + \mu_bP_a + \mu_b\mu_ac); \quad X_{chs} = \mu_acP_b / (\mu_acP_b + \mu_bP_a + \mu_b\mu_ac) \]

Similarly one can derive the limiting condition in the presence of A-AND-B type regulation of promoter C by the dimer of the protein products of genes \( A \) and \( B \). In such conditions the differential equation associated with \( P_c \) in Eqs 9 will be modified as follows.

\[ dP_c/d\tau = \Psi_{+} (X_{cys}) - P_c; \quad X_{cys} = P_cP_b / (P_aP_b + L_{ab\mu_{bc}}) \]

Here one should note that the dynamical variables \( P_a, P_b, \) and \( P_c \) also represent the efficiency of TF genes \( A, B \) and \( C \) in raising their protein levels toward their unregulated steady state values \( \psi_h = k_{mih}k_{ph}/\gamma_h \gamma_{ph} \) when \( h = a, b, c \) even in the presence of various types of positive/negative type regulation on their promoters. The maximum achievable steady state values of \( P_h \) will be \( \psi_h = 1 \). In case of PPP type A-OR-B FFL, \( P_a \) and \( P_b \) influence the change of \( P_c \) in an additive way. Whereas in case of P-P type A-AND-B FFL \( P_a \) and \( P_b \) influence the rate of change of \( P_c \) in a multiplicative way. As a result of this multiplicative effect, P-P type FFL can effectively filter out the short/transient pulses of signals that are introduced at the promoter of TF gene \( A \). When TF proteins \( A \) and \( B \) decay much faster than TF protein \( C \), then we find that as \( \rho_a \) and \( \rho_b \) tend toward zero. Under such conditions, the expression for \( X_{cys} \) in Eqn 10 can be written as follows.

\[ X_{cys} = X_{cys}(\tau) = \chi(t) \chi(t) / (\chi(t) \chi(t) + (\mu_ab + \chi(t))L_{ab\mu_{bc}}) \]

Eqn 11 suggests that when the binding parameters \( \mu_{ab}, \mu_{bc}, \mu_{ac} \) are much lesser than one and the promoter of gene \( A \) is turned on for sufficiently longer time periods then the TF gene \( A \) will be turned on to its maximum expression level since \( X_{ab} \sim 1 \) under such conditions. Similar to Eq 10 when \( \rho_a \) and \( \rho_b \) tend toward zero, then we can write the expression for the occupancy level of the promoter \( C \) by TF proteins of genes \( A \) or \( B \) in case of PPP type A-OR-B FFL as follows.

\[ X_{cs} = (\mu_b\chi(t)/(\mu_ab + \chi(t)) + \mu_ac\chi(t)) / (\mu_ac\chi(t) + (\mu_ab + \chi(t)) (\mu_b\chi(t) + \mu_b\mu_ac)) \]

This equation suggests that when the binding parameters \( \mu_{bc}, \mu_{ac} \) are much lesser than one and the promoter of gene \( A \) is turned on for sufficiently longer time periods, then the TF gene \( C \) will be turned on to its maximum expression level since \( X_{cs} \sim 1 \) under such conditions. Using Eqs 11 and 12 one can derive the steady-state values \( (P_a, P_b, P_c) \) of scaled protein concentrations \( P_a, P_b, \) and \( P_c \) associated with Eqs 9 as follows.

\[ P_{as} = \chi(t); \quad P_{bs} = \Omega_+ (X_{bas}); \quad P_{cs} = \zeta_{ab+bc} (X_{cas}X_{chs}) \]
Stochastic Analysis

The set of Chemical Langevin equations (CLE) associated with the expression of TF genes A, B and C within the various types of A-OR-B type FFLs can be written as follows [26–28].

**TF A:**

\[
\begin{align*}
\frac{dM_{A}}{d\tau} &= Z(\tau) - M_{A} + \sqrt{Z(\tau)} \Gamma_{i}^{m_{A,0}} - \sqrt{M_{A}} \Gamma_{i}^{m_{A,t}} \\
\frac{dP_{A}}{d\tau} &= M_{A} - P_{A} - \sigma_{ab}(P_{A}(1 - X_{B}) - \mu_{ab}X_{A}) - P_{A}Z_{i}^{tbc}
\end{align*}
\]  

\[\tag{16}
\]

**TF B:**

\[
\begin{align*}
\frac{dX_{B}}{d\tau} &= P_{A}(1 - X_{B}) - \mu_{ab}X_{B} + Z_{i}^{tba} \\
\frac{dM_{B}}{d\tau} &= \Omega_{+}(X_{B}) - M_{B} + \sqrt{\Omega_{+}(X_{B})} \Gamma_{i}^{m_{B,0}} - \sqrt{M_{B}} \Gamma_{i}^{m_{B,t}} \\
\frac{dP_{B}}{d\tau} &= M_{B} - P_{B} + \sqrt{M_{B}} \Gamma_{i}^{m_{B,t}} - \sqrt{P_{B}} \Gamma_{i}^{m_{B,t}} - \sigma_{bc}Z_{i}^{tcb}
\end{align*}
\]  

\[\tag{17}
\]

**TF C:**

\[
\begin{align*}
\frac{dX_{C}}{d\tau} &= P_{A}(1 - X_{C}) - X_{B} - \mu_{bc}X_{B} + Z_{i}^{tca} \\
\frac{dX_{B}}{d\tau} &= P_{B}(1 - X_{B}) - \mu_{bc}X_{B} + Z_{i}^{tcb} \\
\frac{dM_{C}}{d\tau} &= \zeta_{a} \pm \zeta_{b} \cdot (X_{B} - X_{B}) - M_{C} + \sqrt{M_{C}} \Gamma_{i}^{m_{C,t}} \\
&+ \sqrt{\zeta_{a} \pm \zeta_{b} \cdot (X_{B} - X_{B})} \Gamma_{i}^{m_{C,t}} - \sqrt{M_{C}} \Gamma_{i}^{m_{C,t}} \\
\frac{dP_{C}}{d\tau} &= M_{C} - P_{C} + \sqrt{M_{C}} \Gamma_{i}^{m_{C,t}} - \sqrt{P_{C}} \Gamma_{i}^{m_{C,t}} + \rho_{c} \cdot Z_{i}^{tca}
\end{align*}
\]  

\[\tag{18}
\]

Eqs 16–18 suggest the presence of non-zero temporal correlations among the set of concentration variables \(\langle X_{B}P_{A} \rangle, \langle X_{C}P_{A} \rangle\) and \(\langle X_{B}P_{B} \rangle\). The characteristic correlation times associated with these pairs of variables will be much lesser than that of the timescales associated with the synthesis and decay of the respective protein products. This follows from the fact that the timescale associated with the promoter state fluctuations is well separated from the timescale associated with the synthesis and decay of the TF proteins. Here one should note that \(\rho_{c} = 1\) by definition. In this equation, various types of \(\zeta\) parameters are defined as follows.

\[
\begin{align*}
Z_{i}^{tba} &= \sqrt{P_{A}(1 - X_{B})} \Gamma_{i}^{m_{B,0}} - \sqrt{\mu_{ab}X_{B}} \Gamma_{i}^{m_{B,t}} \\
Z_{i}^{tca} &= \sqrt{P_{A}(1 - X_{C}) - X_{B}} \Gamma_{i}^{m_{C,0}} - \sqrt{\mu_{bc}X_{B}} \Gamma_{i}^{m_{C,t}} \\
Z_{i}^{tnc} &= \sqrt{M_{A}} \Gamma_{i}^{m_{C,0}} - \sqrt{M_{A}} \Gamma_{i}^{m_{C,t}} - \sigma_{ab}Z_{i}^{tba} - \sigma_{bc}Z_{i}^{tcb} \\
Z_{i}^{tcb} &= \sqrt{P_{B}(1 - X_{C}) - X_{B}} \Gamma_{i}^{m_{C,0}} - \sqrt{\mu_{bc}X_{B}} \Gamma_{i}^{m_{C,t}} \\
Z_{i}^{tbc} &= \sqrt{P_{A}(1 - X_{C}) - X_{B}} \Gamma_{i}^{m_{C,0}} - \sqrt{\mu_{ab}X_{B}} \Gamma_{i}^{m_{C,t}} \\
\end{align*}
\]  

\[\tag{19}
\]

In Eqs 16–19, the term \(\zeta\) is the dimensionless delta-correlated Gaussian white noise with the following mean and variance properties.

\[
\begin{align*}
\Gamma_{t}^{i} &= \sqrt{\lambda_{ab} \xi_{i}(\tau - \tau'); \lambda_{ab} = (\sigma_{ab} \sigma_{bc})^{-1}}; \theta_{ab} = (\mu_{ab} \sigma_{bc})^{-1} \\
\langle \Gamma_{t}^{i} \Gamma_{t}^{j} \rangle &= \lambda_{ab} \delta(\tau - \tau') \\
\langle \Gamma_{t}^{i} \rangle &= 0; k, u = m, p; v = a, b, c; i = o, t; \\
\langle \Gamma_{t}^{i} \Gamma_{t}^{j} \rangle &= 0; u \neq k; v \neq h; i \neq j \\
\langle \Gamma_{t}^{i} \Gamma_{t}^{j} \Gamma_{t}^{k} \rangle &= 0; k = b, c; h = a, b; \\
\langle \Gamma_{t}^{i} \Gamma_{t}^{j} \Gamma_{t}^{k} \rangle &= 0; u \neq k; v \neq h; i \neq j \\
\langle \Gamma_{t}^{i} \Gamma_{t}^{j} \Gamma_{t}^{k} \rangle &= \lambda_{ab} \delta(\tau - \tau'); \lambda_{ab} = (\mu_{ab} \sigma_{bc})^{-1}
\end{align*}
\]  

Equations 16–19 in case of A-AND-B type FFLs there is an additional source of fluctuations arises from the dimerization reaction between the TF proteins A and B. The modified equations in Eqs 16–18 in the presence of A-AND-B type logic can be written as follows.

\[
\begin{align*}
\varepsilon_{ab} dY_{ab}/d\tau &= P_{A}P_{B} - L_{ab}Y_{ab} - \sigma_{y}(Y_{ab}(1 - X_{B}) - \mu_{y}X_{C}) \\
&+ \sigma_{y}Z_{i}^{t+y} + Z_{i}^{t+y} \\
\rho_{c} dP_{A}/d\tau &= M_{A} - P_{A} - \sigma_{ab}(P_{A}P_{B} - L_{ab}Y_{ab}) + \sqrt{M_{A}} \Gamma_{i}^{m_{A,0}} \\
&- \sqrt{P_{A}} \Gamma_{i}^{m_{A,t}} - \sigma_{ab}Z_{i}^{t+y} \\
\rho_{c} dP_{B}/d\tau &= M_{B} - P_{B} - \sigma_{bc}(P_{A}P_{B} - L_{ab}Y_{ab}) + \sqrt{M_{B}} \Gamma_{i}^{m_{B,0}} \\
&- \sqrt{P_{B}} \Gamma_{i}^{m_{B,t}} - \sigma_{bc}Z_{i}^{t+y} \\
\varepsilon_{C} dX_{C}/d\tau &= Y_{ab}(1 - X_{B}) - \mu_{y}X_{C} + Z_{i}^{t+y} \\
\varepsilon_{C} dM_{C}/d\tau &= \Psi_{xy}(X_{C}) - M_{C} + \sqrt{\Psi_{xy}(X_{C})} \Gamma_{i}^{m_{C,0}} \\
&- \sqrt{M_{C}} \Gamma_{i}^{m_{C,t}}
\end{align*}
\]  

\[\tag{20}
\]

In this equation, various types of \(\zeta\) parameters and the Gaussian noise terms \(\Gamma\) associated with the dimerization dynamics are defined as follows.

\[
\begin{align*}
Z_{i}^{t+y} &= \sqrt{Y_{ab}(1 - X_{B})} \Gamma_{i}^{m_{y,0}} - \sqrt{\mu_{y}X_{C}} \Gamma_{i}^{m_{y,t}} \\
Z_{i}^{t+y} &= \sqrt{P_{A}P_{B}} \Gamma_{i}^{m_{y,0}} - \lambda_{ab} \delta(\tau - \tau') \\
\Gamma_{i}^{m_{y,t}} &= \sqrt{\lambda_{ab} \xi_{i}(\tau - \tau')}; \lambda_{ab} = (\mu_{ab} \sigma_{bc})^{-1}; \langle \Gamma_{t}^{i} \rangle = 0 \\
\langle \Gamma_{t}^{i} \Gamma_{t}^{j} \rangle &= \lambda_{ab} \delta(\tau - \tau'); i = o, t \\
\langle \Gamma_{t}^{i} \Gamma_{t}^{j} \Gamma_{t}^{k} \rangle &= \lambda_{xy} \delta(\tau - \tau')
\end{align*}
\]  

\[\tag{22}
\]

**Discussion**

Response times of various type of FFLs are strongly dependent on the sets of parameters \(\mu, \nu, w\) and weakly on \(\sigma\). The sets of
parameters $\bar{\psi}$ and $\bar{\sigma}$ characterize the temporal coupling between the mRNA/protein dynamics and the binding-unbinding dynamics of protein molecules with the promoter. The ordinary perturbation parameter $\bar{\sigma}$ characterizes the strength of temporal coupling between the protein level dynamics with the promoter state occupancies. The variable $\bar{\psi}$ represents the ratios of the lifetimes of various mRNAs to lifetime of protein of TF gene C and reflects the coupling between the mRNA and protein degradation dynamics. The variable $\bar{\mu}$, the set of dissociation constants is inversely proportional to the TFs/promoters binding affinity and characterizes the strength of various regulatory connections. The parameter set $\bar{\sigma}$ does not affect the rise-time of the TF gene C significantly since the change in the number of TF protein molecules due to binding-unbinding at various promoters is negligible. In contrast, $\bar{\psi}$ significantly affects the response times since the effect of varying $\bar{\psi}$ is indirectly amplified through the corresponding mRNA dynamics. Furthermore, an increase in $\bar{\psi}$ would decrease the rate at which promoter state occupancies shift towards saturation (in positive type regulation) or free form (in negatively regulated promoter). Decrease in $\bar{\psi}$ will be generally in the order of $10^{-3}$-10$^{-10}$. Here $\bar{\sigma}$ represents the direct coupling between the promoter state occupancies and the rate of synthesis and degradation of TF proteins. Variation in $\bar{\sigma}$ does not influence the response times much since the number of protein molecules associated with the binding-unbinding events are much less compared to the steady state values.

The dependency of the response times on the parameters $\bar{\psi}$ seems to be strongly influenced by the set of binding parameters $\bar{\mu}$. Under weak binding conditions such as $\bar{\mu} \geq 1$, we find that each of FFLs under consideration show different type of variations as $\bar{\psi}$ changes. Figure 2A suggests that the response times of various FFLs under weak binding conditions seems to be in the descending order of P-P, N-P, PPP, NNP, PNP, PPN, N-N, P-N, NNN and PNN. We further observe that almost all the FFLs show similar type of variation in response time with respect to changes in $\bar{\psi}$ in the dynamic range $\bar{\psi} \in (0.1, 1)$. From Figures 2B, 2C and 2D we find that this scenario significantly changes as the binding strength increases. Under strong binding conditions such as $\bar{\mu} < 1$, the entire set of FFLs can be approximately categorized into three subgroups (Figure 2D) based on the behavior of overall response times with respect to changes in $\bar{\psi}$ viz. Group I: \{P{P, NNP}, {PPP, NNP, N-P}, {P-P, PNP}\}, Group II: \{NPN, {N-N, NNN}, PNN\} and Group III: P-N. Here the order of response times of various subgroups is Group I > Group II > Group III. One can write the segregation pattern in the standard terminology of FFLs as \{(I3, C4), {C1, 14, N-P}\}, \{(P-P, I1), \}, \{(C2, {N-N, I2}, C3), P-N\}. Though Group II and Group III type FFLs possess lesser response times than Group I their response times increase almost linearly upon an increase in $\bar{\psi}$ over the entire range of investigation on a log-log scale.

Within Group I we find three different subsets of FFLs having similar type of response times and their response times are in the order as \{I3, C4\} \{C1, 14, N-P\} \{P-P, I1\}. As we have pointed out in the introduction section, a FFL will be an efficient feedforward logic on overall range of $\bar{\psi}$ since response times of I3 and C4 type FFLs are higher than the generation time of the cell and (2) also it is less sensitive to the signal at the promoter of TF gene A is reasonably low or close to the generation time of the cell and (2) also it is less sensitive to the change in $\bar{\psi}$ over the physiological dynamic range (0.1, 1). Those FFLs which satisfy these two criteria will be the efficient ones and naturally selected. Upon applying these two criteria on the subsets of Group I type FFLs, we find that C1 with both A-AND-B and A-OR-B gated logics and I1 are the preferred FFLs on overall range of $\bar{\psi}$ since response times of I3 and C4 type FFLs are higher than the generation time of the cell. We find from Figure 2D that the increase in the response times is <200% for Group I type FFLs upon increasing $\bar{\psi}$ from 0.1 to 1 whereas it is >400% for Group II and III type FFLs. These results agree well with the earlier observations on the abundance of various FFLs in E. coli [10,13] and yeast [10,14]. Though Group II and III FFLs possess less response times, these FFLs are not selected by nature since the same speeding-up functionality can be achieved through a much simpler negative auto regulatory loops associated with the TF gene C [8–9].

Figures 2E and 2F suggest that the members of Groups I-III are not robust against variation of the parameter set $\bar{\psi}$ under strong binding conditions. Results show that the P-P type FFL will be moved from third subset of Group I to the second one upon an order of increase in $\bar{\psi}$. Further we find that the FFLs such as NPP, P-P, NNP and N-P behave in a similar way with respect to the changes in $\bar{\psi}$ at higher values of $\bar{\psi}$. The segregation pattern of
Figure 2. Dependency of response times on various sets of system parameters viz. $w = (w_\alpha, w_\beta, w_\gamma), \mu = (\mu_{ab}, \mu_{ac}, \mu_{bc}, \mu_{yc}),$ and $\gamma = (\gamma_{ab}, \gamma_{yc}, \gamma_{cl}, \gamma_e)$. Response times are expressed in terms of number of generation times. A. Dependency of response times of various types of FFLs under weak binding conditions. Here the simulation setting are $\mu = 1, \nu = 0.0003, \sigma = 4, \Delta t = 5 \times 10^{-6}$, the total simulation time was set to $T = 25$. 

B. 

C. 

D. 

E. 

F.
various FFLs based on the behavior of the response times over changes in $w$ under such conditions seems to be $\{C4, I3, \{C1, P-P, N-P, I4\}, I1\}, \{C2, \{N-N, I2\}, C3\}, P-N$. Since the value of $v$ is inversely proportional to the speed at which the promoter state reaches the steady state, an increase in $v$ would decrease the transcription and translational rates. This will in turn increase the response times of positively regulated promoters and decrease the response times of negatively regulated promoters. Upon considering all these results together one can conclude that the FFLs of type A, B, P-P, P-P and PNP are robust against changes in both $v$ and $w$. This also could be one of the reasons [14] associated with the observation [10] that these particular coherent C1 type FFLs with both OR and AND-logics are more abundant in nature than the other types. These results are summarized in Table 3.

When the binding parameters $(\mu_{ab}, \mu_{ac}, L_{ab})$ are much less than one, then we find from Eq 11 that the TF gene $C$ will be turned on toward its maximum expression level. This means that strong binding conditions are required for the following types of molecular interactions to achieve the maximum production of TF protein $C$ viz. (1) protein-protein interactions between $A$ and $B$, (2) binding of the dimer $A-B$ at the promoter of $C$ and (3) binding of protein $A$ at the promoter of $B$. Conditions (1–3) also suggest that for an efficient filtering activity of the P-P type A-AND-B FFL against transient input signals at the promoter of the TF gene $A$, the inequality conditions $(\mu_{ab}, \mu_{ac}, L_{ab}) >> 1$ are necessary which in turn will decrease the maximum achievable steady state value of protein $C$ as shown in Figure 3. In other words there exists a critical value of $\tilde{\mu} = \mu_1$ for a given signal with a pulse width $\theta$ above which the response of gene $C$ for a pulse of signal at the promoter of gene $A$ will be practically zero. In Figure 3, this critical value occurs at $\tilde{\mu}_1 = 1$ for a pulse width of $\theta = 0.3$. On the other hand, for a given value of $\mu$ there exist a cutoff pulse width $\theta = \theta_0$ below which the response of TF gene $C$ is practically zero which is demonstrated in Figure 4. From Figure 4 we find that there also exists a transition region in which the TF gene $C$ responds to the width of the input pulse that is given at the promoter of $A$ in a graded manner. From Eq 10 we find that

$$P_t = e^{-t}$$

where $\tau = H(t) - H(t - \theta)$, then we can conclude that the TF gene $C$ will respond to the variation of width $\theta$ of the activation signal in a graded manner in the strong binding limit as $(\mu_{ab}, \mu_{ac}, L_{ab})$ tend toward zero. The maximum achievable $P_t$ will increase proportional to the width of the activation signal. Whereas in the weak binding limit as the binding parameters $(\mu_{ab}, \mu_{ac}, L_{ab})$ tend toward infinity, the response of TF gene $C$ with respect to the pulse width $\theta$ seems to be approximately a sharp type. Under such conditions there exists a critical pulse width $(\theta_0)$ with a sharp transition region above which the maximum achievable $P_t$ will be $P_t = 1$ and below which the maximum achievable $P_t$ will be $P_t = 0$. It seems that the parameters $(\theta_{ab}, \theta_{ac}, L_{ab})$ associated with various binding events need to be fine-tuned to achieve both high efficiency in the filtering activity as well as maximum possible steady-state values of protein $P_t$ upon inducing the promoter of gene $A$ by a persistent signal. Results from stochastic simulations at various values of $v$ and $w$ are shown in Figure 5 and summarized in Table 4. These results suggest that the coefficient of variation in the response times associated with various FFLs under strong

### Table 3. Summary of segregation patterns at weak and strong binding conditions.

| Parameters | Segregation pattern of response times in $w_0$ (0.001, 10) | Abundances of various types of FFLs as in Ref [10] | Source |
|------------|----------------------------------------------------------|--------------------------------------------------|--------|
| Condition I | P-P > N-P > C1 > I4 > C4 > I3 > I2 > N-N > P-P > C2 > C3 | C1 > I1 > C3 > C2 > (C4, I3, I4) | Prokaryotes |
| Condition II | (I3, C4) > (C1, I4, N-P) > (P-P, I1), (C2) > (N-N, I2, C3), P-N | C1 > I1 > C3 > I2 > C2 > (N-N, I2, C3), P-N | Eukaryotes (yeast) |
| Condition III | (C14 > I3) > (C1, P-P, N-P, I4) > I1, (C2 > (N-N, I2, C3), P-N | C1 > I1 > C3 > I2 > C2 > (N-N, I2, C3), P-N | Eukaryotes (yeast) |
| Condition IV | (C14 > I3) > (C1, P-P, N-P, I4) > I1, (C2 > (N-N, I2, C3), P-N, I1 | I1 will be much lower than C1 type. | Higher eukaryotes |

Note: This table summarizes the behavior of various FFLs under strong and weak binding conditions as well as fast and slow promoter-state dynamics. Under weak binding conditions ($\mu_1 = 1$) each FFL behaves differently from each other with respect to changes in $w$. Here the settings for Condition I: weak binding and fast promoter state dynamics ($v = 0.0003, \mu = 0.0003$). Condition II: strong binding and fast promoter state dynamics ($v = 0.0003, \mu = 0.0003$). Condition III: strong binding and slow promoter state dynamics ($v = 0.0003, \mu = 0.0001$). Condition IV: strong binding and slow promoter state dynamics ($v = 0.03, \mu = 0.0001$). Under strong binding conditions $\mu > 0.0001$, the entire set of FFLs segregates approximately into three subgroups I, II and III. Here P-P (C1 type FFL with AND type logic on TF gene $C$ and I1 behaves similarly and therefore the advantages of I1 type FFL whose response time is lower than the generation time will be shared by P-P type which will be added up to the C1 type FFL with OR type gated logic to TF gene $C$. However this pattern seems to be only weakly dependent on $v$. When $v$ increases as in case of eukaryotic cell, then P-P behaves similar to C1 type and I1 type FFL will have the entire advantage of having lower response times than other subgroups of Group I. As a result, I1 type FFL will be more abundant in eukaryotes than prokaryotes. All these results are not dependent on changes in $w$ (0.1, 10), increasing $\nu$ beyond 0.03 or decreasing below 0.0003. One should note that the physiological value of $\nu$ in prokaryotes will be $\sigma < 4$. The overall response time of first subgroup (I3 and C4) of Group I is higher than the generation time of the cell. The response times of the second subgroup (C1, I4 and N-P) are closer to the generation time whereas the third subgroup possesses lesser response times than the others. The response times of the FFLs in Group I are more robust against changes in $w$ (0.1, 10) over the physiological values than Group II and III. Comparison with the relative abundances of naturally occurring FFLs, one can conclude that those FFLs are naturally selected when their response times are (a) robust against changes in $w$ (0.1, 10) and (b) closer to or lower than the generation time.

doi:10.1371/journal.pone.0041027.t003
binding conditions (μ < 0.001) are robust against changes in w, as well as v. When v > 0.003 (as in case of prokaryotes and yeast), then based on the overall coefficient variation in the response times all the FFLs can be categorized into at least three different groups viz. \{I_1, P-N\} > \{C_3, N-N, C_2, I_2\} > \{C_1, C_4, I_3, P-P, N-P, I_4\}. When v > 0.003 (as in case of higher eukaryotes such as human and plants), then the segregation pattern associated with the overall coefficient of variation in the response time of the first group seems to be >100% and in the second group it is 1% and in the third/fourth group of FFLs it is 0.1%. The coherent C1 FFL with both OR/AND type logic shows the least amount of variation in the response times among all the FFLs under consideration.

Earlier results ([10], supplementary materials) based on the analysis of literature-based databases of experimentally verified direct transcription interactions for *E. coli* [6] suggested a distribution pattern of various FFLs as 83% coherent (out of which 80% were C1 type, 11% C3, 6% C2 and 3% C4) and 17% incoherent (out of which 72% were I1, 14% I2 and 14% I4) types. This data suggested an overall distribution of various FFLs in the transcription factor networks of bacteria as (67% C1, 12% I1, 10% C3, 5% C2, 1% C4, 1% I3 and 1% I4). The overall distribution of various FFLs in the TF networks of yeast is (47% C1, 30% I1, 9% C2, 5% I2 and 1% I3). Comparison of the order of preferences of various FFLs in bacteria (C1 > I1 > C3 > C2 > C4, I3, I4) and yeast (C1 > I1 > C2 > I2 > I3) with the Group I, II and III types of FFLs, one can conclude that those FFLs of Group I (C1, I1) whose response times are close to or less than the generation time of the cell apart from the robustness against changes in w are more preferably selected. The next set of preferable FFLs (C2, C3 and I2) is from Group II. The set of FFLs (I3, I4 and C4) are the least preferable ones. Although I3 and C4 fall within Group I, their response times are much higher than the generation time of the cell apart from the robustness against changes in w. Here one should note that the value of the parameter set v = 0.0003 that is used in our simulations (Fig. 2A-D) is derived for a prokaryote organism whose cell volume is ~10^{-18} m^3. With this setting we find from our simulations that the subset (P-P and I1) behaves in a similar way within Group I which is evident from the segregation pattern of various FFLs at strong binding conditions \{I_3, C_4\}, \{C_1, H, N-P\}, \{P-P, I_1\}, \{C_2, \{N-N, I_2\}, C_3\}, P-N.

Figure 3. Dependency of filtering efficiency of P-P type (C1 type with AND-logic) FFL on the set of binding parameters μ. Here the general settings are: w = 0.12, v = 0.0003, θ = 4, T = 25 generation times, and μ was varied as (0.1, 0.2, 0.3, 0.4, 0.5, 1, 3 and 5). A. Input signal at the promoter of TF gene A. This has one short rectangular pulse with a width of θ = 0.3 and a large one with θ = 8 all are measured in terms of number of generation times. B. TF gene A responds similarly for all values of μ to both the signals irrespective of the binding strengths. The response seems to be proportional to the pulse width θ without any delay. C. As binding strength increase, the response of B also increases proportionately. D. There exists a cutoff value of μ above which the expression level of TF gene C is practically zero. With the current settings, this cutoff seems to occur at μ = 1. Arrow shows the increasing direction of μ.

doi:10.1371/journal.pone.0041027.g003
type FFL in eukaryotes than the prokaryotic TF networks. Further increase in \( \varphi \) as in case of nucleus of human cell \(( = 0.03)\) shift the I1 type FFL from Group I to Group III. This result predicts that I1 type FFLs will be much lesser in abundance than the C1 type in the TF network of human and other higher animals and plants. One also should note that we are still not able to explain the relative scarcity of I4 and N-P type FFLs even though they fall well within Group I. As pointed out by Magnan and Alon in reference [10], the relative scarcity could be due to the reduced functionality of type 3 and type 4 FFLs with AND gated logic (here it is applicable to N-P) since they respond at most one of the triggering signals given at the promoters of both the genes A and B.

We have considered the sets of similar parameters \( (w, \varphi, \varphi, \sigma, \rho) \) associated with TF genes A, B and C as variable units rather than individual ones for all these calculations and simulations. The TF network of an organism consists of several FFL motifs and each of these FFLs is constituted with different subsets of a pool of TF genes. Each subset of TFs can be represented as points in the parametric space of our model. Upon considering the entire pool of TF genes, one should note that each of the members of these sets of parameters can take a spectrum of values with a probability distribution with definite mean and variance. There are two types of variability of parameters among the TF genes of FFLs viz. variation of the parameters of TF gene A/B/C across various FFLs found in the entire TF network and variation of the parameters of TF genes A, B, C within a given FFL. Here we have assumed that both of them are approximately the same. The main results of our analysis will not be affected much due to the second type of variation since we have used the mean values of the parameters associated with the entire set of TF genes of an organism to represent genes A, B and C of various types FFLs.

**Methods**

We use the following Euler type iterative numerical scheme to integrate the deterministic differential Eqs 1–3 associated with various A-OR/AND-B type FFLs in the dimensionless time and concentrations space.

**TF A:**

\[
M_{a,n+1} = M_{a,n} + w_a^{-1}(P_{a,n}(1 - X_{b,a}) - \mu_a X_{b,a}) \Delta \tau
\]

\[
P_{a,n+1} = P_{a,n} + \rho_a^{-1}
\]

\[
\left(M_{a,n} - P_{a,n} - \sigma_a(P_{a,n}(1 - X_{b,a}) - \mu_a X_{b,a}) - \sigma_a(P_{a,n}(1 - X_{c,a} - X_{c,b}) - \mu_a X_{c,a})\right) \Delta \tau
\]

**TF B:**

\[
X_{b,a,n+1} = X_{b,a,n} + w_b^{-1}(P_{b,n}(1 - X_{b,a}) - \mu_b X_{b,a}) \Delta \tau
\]

\[
M_{b,n+1} = M_{b,n} + w_b^{-1}(\Omega_2(X_{a,n}) - M_{b,n}) \Delta \tau
\]

\[
P_{b,n+1} = P_{b,n} + \rho_b^{-1}(M_{b,n} - P_{b,n} - \sigma_b(P_{b,n}(1 - X_{a,n} - X_{c,b}) - \mu_b X_{a,b})) \Delta \tau
\]
Figure 5. Coefficient of variation (CV) associated with the fluctuations in the response times of various types of FFLs under strong binding conditions ($\mu = 0.001$). CV was calculated over $10^5$ numbers of stochastic trajectories. The coherent C1 type FFL possesses lower CV of response times than other FFLs whereas I1 type possesses highest CV of response times. A. Here the settings are $\gamma \leq 0.003$ that is applicable to both prokaryotes and eukaryotes such as yeast, $\gamma$ was iterated from 0.001 to 10, and $\sigma = 4$, $T = 25$ generation times. B. Here the settings are $\gamma \geq 0.03$ that is applicable to higher eukaryotes such as human, $\gamma$ was iterated from 0.001 to 10, and $\sigma = 4$, $T = 25$ generation times.

doi:10.1371/journal.pone.0041027.g005
### Table 4. Segregation patterns of CV (response time) at weak and strong binding conditions.

| Parameters | Segregation pattern of CV (response-time) in \( \omega=(0.001, 10) \) | Abundances of various types of FFLs as given in Ref [10] | Source |
|------------|-----------------------------------------------|------------------------------------------------|--------|
| Condition II (\( [I_1, P-N] > [C_3, N-N, C_2, I_2] > [C_1, C_4, I_3, P-P, N-P, I_4] \)) | \( C_1 > I_1 > C_3 > C_2 > (C_4, I_3, I_4) \) | Prokaryotes |
| Condition III (\( [I_1, P-N] > [C_3, N-N, C_2, I_2] > [C_1, P-P, C_4, I_3, N-P, I_4] \)) | \( C_1 > I_1 > C_3 > I_2 > I_3 \) | Eukaryotes (yeast) |
| Condition IV (\( [I_1, P-N] > [C_3, N-N, C_2, I_2] > [C_1, P-P, N-P, I_4] \)) | \( I_1 \) will be lower than \( C_1 \) type. | Higher eukaryotes |

**Note:** This table summarizes the results from the stochastic simulation. Parameter settings for **Condition II:** strong binding and fast promoter state dynamics (\( \phi=0.0003, \sigma=4, \mu=0.001 \)). **Condition III:** strong binding and slow promoter state dynamics (\( \phi=0.003, \sigma=4, \mu=0.001 \)). **Condition IV:** strong binding and slow promoter state dynamics (\( \phi=0.03, \sigma=4, \mu=0.001 \)). Here CV (defined as the ratio standard deviation/mean) represents the coefficient of variation in the response times (the time required to attain half of the steady state value of the transcription factor protein C in the FFLs).

doctoral.13737/journal.pone.0041027.t004

TF C:

\[
X_{a,n+1} = X_{a,n} + v_{a}^{-1} \left( P_{a,n} (1 - X_{a,n} - X_{b,n}) - \mu_{a} X_{a,n} \Delta t \right) \\
X_{b,n+1} = X_{b,n} + v_{b}^{-1} \left( P_{b,n} (1 - X_{a,n} - X_{b,n}) - \mu_{b} X_{b,n} \Delta t \right) \\
M_{c,n+1} = M_{c,n} + w_{c}^{-1} \left( \gamma_{c} X_{a,n} X_{b,n} - M_{c,n} \right) \Delta t \\
P_{c,n+1} = P_{c,n} + (M_{c,n} - P_{c,n}) \Delta t
\]

(25)

To investigate the filtering efficiency of P-P type A-AND-B FFL, we can use the signal function \( \chi(t) = \sum_{m=1}^{2} (H(t-\tau_{m}) - H(t-\tau_{m}-\theta_{m})) \) at the promoter of TF gene A with two well separated rectangular pulses with respectively lower and higher pulse widths. For the purpose of computing the response times we can reset the input function at the promoter of gene A as \( \chi(t) = H(t-T) \) where \( T \) is the total simulation time. The numerical scheme for the modified Eqs 6–8 for \( Y_{a,b} = P_{a}, P_{b}, X_{b} \), and \( M_{c} \) in case of A-AND-B type FFLs can be written as follows.

\[
Y_{a,b,n+1} = Y_{a,b,n} + v_{a,b}^{-1} \left( P_{a,b} (P_{b,n} - L_{a,b} Y_{a,b,n}) \right) \Delta t \\
P_{a,b,n+1} = P_{a,b,n} + \rho_{a,b}^{-1} \left( M_{a,b} - P_{a,b} - \sigma_{a,b} \left( P_{a,b} (1 - X_{a,b,n} - \mu_{a} X_{a,b,n}) \right) \right) \Delta t \\
\sigma_{a,b} \left( P_{a,b} (1 - X_{a,b,n}) \right) \Delta t \\
M_{b,n+1} = M_{b,n} - \sigma_{b} \left( P_{a,b} (P_{b,n} - L_{a,b} Y_{b,n}) \right) \Delta t \\
M_{c,n+1} = M_{c,n} + w_{c}^{-1} \left( \Psi_{c} (X_{a,n} X_{b,n}) - M_{c,n} \right) \Delta t
\]

(26)

The total scaled simulation time \( T \) will be divided into \( N \) equal intervals such that \( \Delta t = T/N \). For simulation we set \( \Delta t = 5 \times 10^{-3} \) and the corresponding \( \Delta t = 0.2s \) for a lifetime 1/\( \gamma_{p} \) ~ 60 mins. The initial conditions are \( X_{a,b,0} = 0, M_{a,b,0} = 0 \) and \( P_{a,b,0} = 0 \) for \( h = a, b, c \) and \( g = b, c \). To compute the response-time of the TF gene C we set an absorbing boundary at \( P_{a}/P_{b} = 1/2 \) where \( P_{b} \) is the steady state value of scaled concentration \( P_{b} \) which in turn depends on the type of FFL under consideration as given in Eq 13. We measure the concentrations in terms of number of molecules inside the cell. Considering a bacterial cell (volume ~ 10^{-12} m^3) [21] we set \( dm = 1, ma = 10^2 \) molecules and \( h_{a,b} \sim 10^3 \) molecules [22,23] where \( h_{a,b} \sim 10^3 \) molecules. Concentration of a single TF molecule inside a bacterial cell will be ~2 nM. Concentration of a single TF molecule inside the nucleus of yeast cell [24,25] will be ~200 pM and inside the nucleus of human cell it is ~20 pM. We measure the timescales in terms of the lifetime of the TF protein C by dividing \( t \) by \( \gamma_{p} \). Since the dynamics of binding-unbinding of the transcription factors with the respective promoters is a typical diffusion-controlled site-specific DNA-protein interaction, under in vivo conditions of bacterial cell we find that \( k_{d,b} \sim 10^{-3} \) molecules~1^{-1} [22–24]. Here we have assumed an in vivo three dimensional diffusion controlled collision rate ~10^{-6} M^{-1}s^{-1}. In case of nucleus of yeast cell we find \( k_{d,b} \sim 10^{-4} \) molecules~1^{-1} and in case of nucleus of human cell we find \( k_{d,b} \sim 10^{-5} \) molecules~1^{-1}. Here the TF protein \( k \sim (a, b) \) binds with the promoters of \( h \sim (b, c) \). It will be a complicated task to explore the entire parametric space. To simplify the analysis further we can consider the sets of parameters viz. \( \phi = (\phi_{bb}, \phi_{ab}, \phi_{ba}, \phi_{aa}), \sigma = (\sigma_{bb}, \sigma_{ab}, \sigma_{ba}, \sigma_{aa}), \gamma_{p} \) as the parametric units of our numerical simulations. When all the TF proteins \( A, B \) and \( C \) decay with similar decay rate constants then we find that \( \rho_{a,b} = 1 \). Using the steady state values of various proteins and mRNA numbers in the absence of any regulation one obtains \( \phi \sim 3 \times 10^{-4} \sim 0.003 \) for nucleus of yeast cell and \( \sim 0.03 \) for nucleus of human cell and \( \sigma \sim 4 \sim 0.4 \) for yeast cell nucleus and \( \sim 0.04 \) for nucleus of human cell where we have used a protein decay rate \( \gamma_{p} \sim 3 \times 10^{-4} \) [for a protein lifetime of ~60 min] to transform the real time variables to \( \tau \) space variables. The dependency of various properties of TF gene C such as response-time, delay with respect to the induction signal at the promoter of gene A and maximum achievable \( P_{a} \) on the parameter sets \( \mu, \nu, \sigma \) and \( \omega \) were explored. The values of \( \phi_{a,b} \) can vary across the spectrum of genes from \( \phi_{a,b} \sim 10^{-3} \) in case of prokaryotes to \( \phi_{a,b} \sim 1 \) in case of eukaryotes [17–19]. Together with all these values we also iterated the parameter set \( \phi \) inside the range (0.001, 10) with a step size of \( \Delta \phi \sim 0.001 \). For further exploratory purposes we considered four different binding conditions from low to high viz. \( \phi = (0.001, 0.01, 0.1, 1) \) and three different values of \( \sigma = (0.0003, 0.0003, 0.03) \). Similar to Eqs 21–24 we use Euler type numerical scheme to integrate the Chemical Langevin Eqs 16–20 where we replace \( \Gamma \) with Gaussian distributed random numbers with zero mean and unit variance. Assuming an overall in vivo steady state mRNA levels as \( m_{a,b} \sim 10^2 \) and proteins levels as \( p_{a,b} \sim 10 \) where \( h = a, b, c \),
measured in number of molecules) we find that $\lambda_{nh} \sim 10^{-2}/\omega_h$ molecule$^{-1}$ and $\lambda_{ph} \sim 10^{-3}/\rho_h$ molecule$^{-1}$. Further we also find that $\lambda_{nh} \sim 1$ s, $\lambda_{nh} \sim 10^{-3}$ molecule$^{-1}$ s $(k = b, c$ and $h = a, b)$ and $\lambda_{ph} \sim 10^{-2}$ molecule$^{-1}$ s. We use the reflecting boundaries $(0, 1)$ for the scaled concentration variables $\{P_h, M_h, I_{nh}, X_{nh}\}$ where $n = b, c$ and $k = a, b$ and the absorbing boundary condition $P_h = \frac{1}{2}$ to compute the mean first passage time (response-time) from the stochastic simulations. Averaging was done over $10^5$ trajectories at each $w_i$ value and the coefficient of variation of response time was computed as $CV$ (response-time) = standard deviation of response-time/mean of response-time.

Conclusions

Feedforward loops (FFLs) consist of three genes which code for three different transcription factors A, B and C where B regulates C and A regulates both B and C. We have developed a detailed model to describe the dynamical behavior of various types of coherent and incoherent FFLs in the transcription factor networks. We considered the deterministic and stochastic dynamics of both promoter-states and mRNAs of various genes coding for the transcription factors associated with the FFL motifs. Detailed analysis showed that the response times of FFLs strongly dependent on the ratios $(\alpha_h = \gamma_p / \gamma_{ph}$ where $h = a, b, c)$ between the lifetimes of mRNAs of A, B and C $(1/\gamma_{nh})$ and the protein of C $(1/\gamma_{ph})$. When the binding of transcription factors A and B with the cis-acting elements of respective promoters is very strong, then we could categorize all the possible FFLs into Group I, II and III based on the dependency of the response times of FFLs on $\alpha_h$. Though the response times of Group I FFLs were higher than II and III, they seem to be less sensitive to the changes in $\alpha_h$ within the naturally occurring dynamic range $(0, 1)$. We have further shown that among the members of the Group I FFLs, the coherent C1 type was more robust against changes in other system parameters which could be one of the reasons why C1 type coherent FFLs are more abundant in nature than the others.

Author Contributions

Conceived and designed the experiments: RM. Performed the experiments: RM. Analyzed the data: RM. Contributed reagents/materials/analysis tools: RM. Wrote the paper: RM.